Insights on Enzymes and Polymers from Molecular Dynamics Simulations: Applications to Dihydrofolate Reductase Complexes and Starburst Dendrimers

Thesis by

Adel Marie Naylor

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted February 1, 1989)

This thesis is dedicated

to my family:

Tom, Titi, Maura, and Joel Naylor

and

Adeline Parente

and to my "adopting" families:

the Baughs, Kantorczyks, and Nollars

To laugh often and love much;
to win the respect of intelligent persons
and the affection of children;
to earn the approbation of honest critics
and to endure the betrayal of false friends;
to appreciate beauty;
to find the best in others;
to give of one's self;
to leave the world a bit better,
whether by a healthy child, a garden patch,
or a redeemed social condition;
to have played and laughed with enthusiasm
and sung with exultation;
to know that even one life has breathed easier
because you have lived —
this is to have succeeded.

Ralph Waldo Emerson

For unto whomsoever much is given,
of him shall be much required;
and to whom men have committed much,
of him they will ask the more.

Luke 12:48

Acknowledgments

The passage in Luke 12:48 holds a great deal of significance for me. I know that much has been given to me...a loving family, cherished friends, intellectual abilities, and a deep faith, to list but a few. To that end, one of the most special (and challenging) gifts I have received is the opportunity to do graduate study at Caltech. A number of individuals contributed significant time, energy, and love to make my California experience a productive and memorable one.

Bill Goddard is an enthusiastic and inspiring research advisor whose love for science is contagious (and tireless!). The privilege of working with a scientist as talented and as insightful as Bill is an experience offered to only a few. I was fortunate that the opportunity was presented to me. I also want to thank Yvonne Goddard for her sacrifices to accommodate the time that Bill spends away from his family to be with his research group.

Additionally, I must thank my "pseudo" second advisor, Steve Benkovic. His continued interest in our DHFR simulations, and encouragement when times were tough, will always be remembered with gratitude. It has indeed been a privilege, and a pleasure, for me to collaborate with Steve and his Penn State associates Carol Fierke, Chris Falzone, and Joe Adams.

Don Tomalia's enthusiasm for the Starburst Dendrimers sparked the simulation studies.

A number of agencies supplied financial support for my endeavors at Caltech. The Department of Energy, Energy Conversion and Utilization Technologies generously funded my graduate research. Triton Bioscience also supplied funds for part of this work. An ARCS Foundation Fellowship assisted in the acquisition of texts and in defraying some living expenses. A Dreyfus Travel Fellowship enabled

me to visit a number of industrial research labs.

Various past and present members of the ever-growing Goddard Research Group have made the hours at Caltech challenging, interesting, and, of course, educational. The "old timers," i.e., Barry Olafson and Marv Goodgame, provided many amusing examples of group folklore through both actions and tales. Emily Carter presented the ultimate challenge; I grew to be a bit stronger in character because of her. Mark Brusich always spoke kind words, performed generous deeds, and offered tremendous amounts of support and encouragement. Kim Schugart helped me to keep an appropriate perspective on graduate school. Barbara Garrison provided friendship, guidance, and was an excellent role model. Janice Peters offered the example of a true Christian. Alan Mathiowetz kept me chuckling with his somewhat infrequent WAG's World cartoon strip and helped to blow off some Pasadena dust with our DDs. Terry Coley always managed to sustain a levelheaded, calm state, no matter what the situation or crisis; lucky for me he had the office next door. Jon Hurley provided great computer help and some amusing anecdotes on life. Ann Miller always offered a cheery good morning and delicious matzo ball soup. Kevin Plaxco brought a biological perspective, and many smiles, to 148 Noyes. Gilles Ohanessian shared his amusing wit and worldly perspective on life and science. Carol Scrivner was a great help in translating WAGscript and providing a quiet, secluded office. And finally, Adria McMillan and Siddharth Dasgupta were always available to supply constant doses of support, encouragement, and love. Life in the Goddard Group would have been significantly less pleasant for me without Adria's hugs and Dr. D.'s comaraderie.

There are also a few special non-group Techers that made the time in Noyes more enjoyable. David Kaisaki is an exceptional friend. He has an uncanny

sense of knowing when to encourage, when to tease, and when to supply Godiva chocolate! Bob (JB) Driscoll kept me smiling with amusing stories and great company. Mary Selman was very encouraging during the job hunt and thesis writing times. Paul Hipes was a great quantum TA and a fine example of how one graduate student should help another.

My involvement in St. Philip's parish kept me sane through some of the more difficult periods of my California passage. Anne Burgard is a valued friend and has helped me to begin to "let go and let God." Heidi Heil and Karen Royer allowed me the privilege of accompanying them on the beginning of their RCIA journeys. Many other RCIAers, especially Judy Nollar, Shirley Kealey, Sharon Kealey, Marty Smith, Kevin and Jana Huerter, Bruce and Kola Popp, Tom and Janet Selenski, broadened my understanding and experience of Christianity. Gary Bauler, a.k.a. Msgr. B., provided exceptional spiritual nourishment and warm friendship. I am very grateful for the many kind gestures, words of encouragement, and loving prayers sent my way from all of my special friends at St. Philip's.

A number of families welcomed me into their lives while I was on the West Coast. Gene and Dora De Blasio were my California grandparents. They gave the Thanksgiving holiday a unique Pennsylvania flair and made Palm Springs get-away weekends special treats. Leo, Beth, and Bill Cowan provided that sense of security associated with "going to Grandma's." I would have felt quite displaced were they not just an hour's drive away! Mary Cowan was more like a big sister than a cousin and John Sadler helped me keep the family dynamics in perspective. Don, Judy, Donald, Annemarie, Jenny, and Joe Nollar made Pasadena home and introduced me to the delights of California nature. Kevin and Jana Huerter allowed me to delight in the joy of their children Alex, Justin,

and Ryan and to form an ever-so-special bond with Justin Quinlan!

Some other new friends entered my life while I was at Caltech. Shirley Whitmore was a fun roommate; she gave me lessons in the art of "blending" frosting and I, in return, taught her how to remove labels from wine bottles. Wally Reiher always answered my questions about BIOGRAF in an honest and direct manner; he offered alot of encouraging words, too. Sharon Kealey is a special gift from the Lord. She generously opened her home, and her heart, to me. Sharon showed me what it takes to be a maid-of-honor and a responsible godparent. She made the last few years of graduate school so much more pleasant. Laurie Carlson made life more cheerful; she was a great walking companion, needlework buddy, and confidante. Laraine Balk was a great job-hunt consultant and I will *never* forget our freeway experience!

My friends from pregraduate school days remain the extra special people that I have grown to love and depend upon. Diane Balk Palguta continues to show me what constitutes true friendship. My life would be so greatly deminished without her presence (thank you Phyllis!). Eileen Ogelesby Hemminger keeps me smiling with her thoughtfulness and her absurdities. Larry, Mary, Matt, Laura, Nick and Julie Baugh are rare and cherished jewels. Words cannot adequately convey just how special each one is to me. Judy, David, Todd, and Amy Kantorczyk are my first adopting family. Their love and encouragement made the transition into adulthood a bit easier. Thank goodness, though, that I'll finish school before Amy starts college!

My Dow friends continued to support my scientific endeavors throughout graduate school. Jack Martin and Stan Strycker encouraged my pursuits. Linda Myers nurtured my dreams and helped me deal with the realities of graduate school. John Daniel and Bert Schuster were great friends. Emily Matthews

remained my "sister-in-Baugh." Bob Broersma was always willing to chat about cardiovascular pharmacology, especially if it involved dopamine! Then there was Bob Vance. Without Bob's gentle harassing I would not have even made application to Caltech, much less accepted the challenge. And, again, without Bob's patient ear and hearty laugh, I probably would not have met the challenge.

And, finally, the greatest supply of support came from my family. My parents, Tom and Titi, my sister Maura, my brother Joel, and my Nanny Adeline, all banded together to shower me with tremendous amounts of love and encouragement. I shudder to think of how different my life would be if it were not for the education they struggled to provide and the values they instilled through their examples. I love each one of you!

Abstract:

Examples are given for the role of simulation and theory in designing artificial biomimetic and biocatalytic systems. Simulations on the β-alanine starburst dendrimer polymers indicate that, for the higher generation systems: (i) $\sim$50% of the surface area is *internal*, and (ii) $\sim$50% of the spheroidal volume is solvent-filled. These studies suggest a design for encapsulating and delivering dopamine to the kidney for cardiovascular therapies. Investigations of the penta-erythritol based polyether starburst dendrimers show that: (i) the later generations lack any internal surface area or volume and (ii) the dense-packed limit for these polymers to be the third generation, consistent with experiment. For Dihydrofolate Reductase (DHFR), the modeling and simulations: (i) explain the high degree of kinetic similarity between two dissimilar forms of DHFR; (ii) indicate why site specific mutation (Phe-31$\rightarrow$Tyr-31, Leu-54$\rightarrow$Ile-54, or Leu-54$\rightarrow$Gly-54) causes a significant change in the catalytic rate; and (iii) suggest modifications to engineer *E. coli*·*Chicken* hybrid proteins capable of reducing folate.

Table of Contents

Overview of the Thesis

The overall goal of our research is to develop theoretical tools useful in designing artificial biocatalytic systems. In the long term, one would envision the development of biocatalysts that could convert methane into more valuable (or more transportable) fuels or could convert syngas ($CO + H_2$) into various fuels and feedstocks. In the short term, we are focusing on systems capable of selective oxygenations or reductions. We hope to develop principles that would allow us to design and develop artificial systems having the selectivity of the cytochrome P-450 enzymes that play a key role in the oxidation of long chain alkanes and the epoxidation of alkenes. Although selective, the various cytochrome P-450 enzymes involve complex assemblies (cofactors, co-enzymes) that would be difficult to reproduce and control in an artificial system. Thus, we would hope to incorporate catalytic control procedures as effective as those in P-450, yet delete the excess baggage special to biological systems (e.g., cofactor, bulk protein necessary to elicit stability and to create substrate binding pockets, etc.)

Consequently, we have focused our research on two aspects of this complex problem: (i) the design of artificial systems that possess precision, controlable sites capable of selectively encapsulating and delivering smaller molecules, and (ii) understanding the detailed structural and chemical aspects of enzyme active sites involved in and affecting the important catalytic processes, i.e., substrate binding, rate of catalysis, product release.

We begin by presenting a chapter written for the ACS Monograph Series on *Impact of Surface and Interfacial Structure on Enzyme Activity.* It gives a concise overview of our work and serves as an excellent introduction to the studies presented, in much greater detail, in sections two and three of this thesis.

The second section of the thesis presents the details of our simulation studies on two classes of starburst dendrimer polymers. This work represents the theoretical component of a collaborative project, including experimental chemists at Dow Chemical under the direction of Donald Tomalia.

The first class of starburst dendrimers, typified by the macromolecules based on an ammonia core and β-alanine monomers, possesses open and extended forms for the early generations (one through three) but adopt more spheroidal structures for generations 5 and above. The higher generation β-alanine dendrimers contain internal hollows, connected by channels, that run the length of the entire polymer assembly. Our studies on these higher generation polymers indicate that: (a) $\sim$50% of the surface area is *internal*, and (b) $\sim$50% of the spheroidal volume is solvent-filled. Such systems provide the opportunity to selectively design unique polymers with internal binding sites. The sequestering of dopamine in a catechol modified starburst dendrimer is included to illustrate the potential utility of these dendritic materials as therapeutic delivery systems.

The second class of starburst dendrimers, typified by the penta-erythritol polyether series, is found to have spherical forms at very early generations. For this type, the later generations lack any internal surface area or volume. Calculations of the amount of surface area available to terminal generation hydroxyl groups indicate the dense-packed limit for these polyether dendrimers to be the third generation, consistent with experiment.

The third section of the thesis focuses on the results of our investigations on complexes of Dihydrofolate Reductase (DHFR). These represent the fruit of a collaborative experimental/theoretical program designed to probe the details of DHFR's catalytic requirements and properties. The experimental component of the project is directed by Professor Stephen J. Benkovic (The Pennsylvania

State University). The theoretical portion of the research project is being done under the supervision of Professor William A. Goddard III at Caltech's Molecular Simulation Facility. The work presented here focuses on three aspects of this collaborative effort.

First, we discuss the analysis of the sequence, structure, and chemical homology of three distinct forms of DHFR (*E. coli* and *L. casei* bacterial sequences and the *Chicken* form). Experimental kinetic studies indicate a remarkable similarity in the free-energy profile for the reduction of dihydrofolate by *E. coli* and *L. casei* DHFR, despite a mere 27% sequence homology (*vide infra*). We use computer graphics and macromolecular docking, combined with homology assignments, to explain this congruence in the catalytic reactivity of the bacterial systems. The analysis is then extended to include the *Chicken* form of the protein, which has a distinctly different reaction profile and is more similar in function to mammalian DHFR. This interspecies comparison provides the opportunity to begin piecing together an understanding of what contributes to the subtle differences in reductive chemistry exhibited by these enzymes.

The second component of the DHFR investigations involves molecular simulation studies of wild-type *E. coli* DHFR binary (with methotrexate inhibitor) and ternary (with dihydrofolate substrate and NADPH cofactor) complexes. These calculations offer insights on the catalytic reduction of dihydrofolate to tetrahydrofolate as well as control studies for computer mutagenesis experiments. A brief discussion of the simulation methods employed in these theoretical investigations of DHFR complexes is also included.

And finally, the third segment focuses on the selected modification of the *E. coli* DHFR using both single-site mutation and protein loop substitutions. The site-directed mutagenesis studies probe the role(s) of specific amino acid

residues in the functioning of *E. coli* DHFR. Segment mutants were constructed to engineer a bacterial DHFR protein with properties more similar to those of the mammalian enzyme (which has the ability to reduce folate to dihydrofolate). These hybrid proteins are based on the bacterial DHFR (from *E. coli*) where a loop, with amino acid sequence identical to the corresponding region of the *Chicken* DHFR, has been inserted into the protein matrix.

Thus, we will show in the ensuing sections that the application of molecular simulation tools provides new insights and deeper understanding of experimental studies. As experience with such simulations accumulates, and as predictions are made on more complex systems amenable to experiment, it will become increasingly feasible to use the theoretical methods on unknown systems.

Section I

Application of Simulation and Theory to Biocatalysis and Biomimetics

The text of this section is a chapter coauthored with William A. Goddard III. It has been submitted to the ACS Monograph Series on *Impact of Surface and Interfacial Structure on Enzyme Activity.*

Application of Simulation and Theory to Biocatalysis and Biomimetics

Adel M. Naylor and William A. Goddard III

Contribution No. 7818 from the
Arthur Amos Noyes Laboratory of Chemical Physics,
California Institute of Technology, Pasadena, California 91125

Examples are given for the role of simulation and theory in biocatalysis and biomimetics. Simulations on the novel Starburst Dendrimer polymers are used to suggest a design for encapsulating and delivering dopamine to the kidney for cardiovascular therapies. For Dihydrofolate Reductase (DHFR), the simulations (i) indicate why a particular mutation (Phe-31 to Tyr-31) causes a significant change in the catalytic rate, (ii) explain the high degree of kinetic similarity between two dissimilar forms of DHFR, and (iii) suggest that the stable form of the enzyme/substrate complex is *not* the reactive one.

A goal of our research efforts is to develop the theoretical tools useful in designing artificial biocatalytic systems for the production of fuels and chemical feedstocks. In the long term, one would envision the development of biocatalysts that could convert methane into more valuable (or more transportable) fuels or to convert syngas ($CO + H_2$) into various fuels and feedstocks. In the short term, we are focusing on systems capable of selective oxygenations or reductions. We would hope to develop artificial systems with the selectivity of cytochrome P-450 enzymes that play a role in selective oxidation of long chain alkanes and selective epoxidation of alkenes (1). Although selective, the various cytochrome P-450 enzymes involve complex assemblies (cofactors, co-enzymes) that would be difficult to reproduce and control in an artificial system (see Figure 1). Thus, we would hope to incorporate catalytic control procedures as effective as those in P-450, yet delete the excess baggage special to biological systems (e.g., cofactor, bulk protein necessary to ellicit stability and to create substrate binding pockets, etc.)

To that end we have focused our research on two aspects of this complex problem: (i) the design of artificial systems that possess precision sites capable of selectively encapsulating and delivering smaller molecules, and (ii) understanding the detailed structural and chemical aspects of enzyme active

Figure 1. Schematic representation of cytochrome P-450 indicating the excess baggage carried by many biological catalysts.

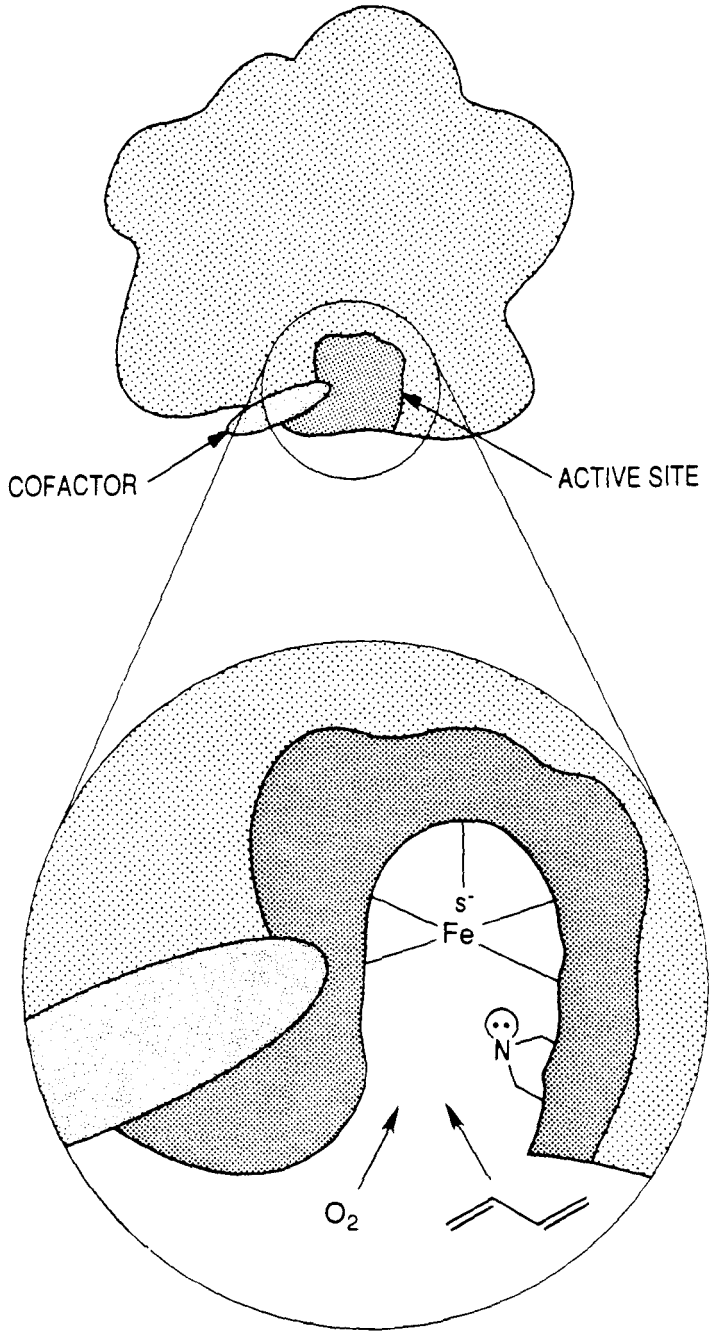


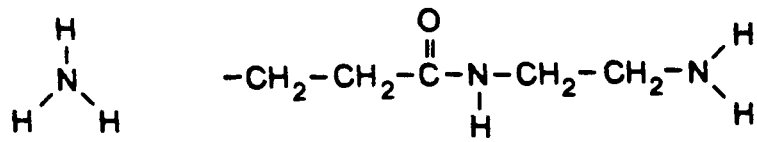
Figure 1.

sites involved in and affecting the important catalytic processes, i.e., substrate binding, product release, rate of catalysis.

Starburst Dendrimers : Molecular Encapsulation

A recent synthetic development has led to a new class of polymers called "starburst dendrimers" (2) that provide the opportunity for design of precise encapsulation strategies. These novel materials (i) start with an initiator unit (termed the "core") that possesses multiple sites for monomer condensation and (ii) use monomer subunits terminating in functional groups that allow for regularized growth and multiple branching at the next generation. With the appropriate use of protecting groups and synthetic strategies, each generation of monomer condensation can be completed before embarking on a new one. Because of the unique chemical requirements for dendrimer formation, these polymers grow in a systematic manner, producing materials with a well-defined number of monomer subunits and a quantized number of terminal groups (Figure 2). With precise topology and chemical properties suitable for the encapsulation of specific target molecules that could be delivered to particular organs (by recognition of surface molecules), one could, in principle, change the monomer from generation to generation, developing onion-like molecules where every layer of the onion has a different chemical composition, thickness, or number of branching sites. Thus, one can imagine designing *and* synthesizing complex polymer structures where (i) internal layers are suitable for encapsulation of the desired molecule, (ii) outer layers are suitable for recognizing binding sites on the organ to which the target molecule is to be delivered, and (iii) intermediate layers protect the target molecule in the blood stream but release them (a) upon binding of the outer layer to recognition sites, (b) upon radiation with external light or charged particles or (c) upon change in acidity or solvent associated with the target organ.

β – Alanine Dendrimers. To illustrate this application, consider the β-alanine dendrimers where the core unit is ammonia (with three condensation sites) and the monomer unit is an amino amide (with two branch sites) as depicted below.



Core Monomer

Experimentally it has been possible to develop complete dendrimers up through the tenth generation. However, no experimental structural data are available (these materials are fractal and do not crystallize). We have been carrying out molecular dynamics simulations on these β-alanine dendrimers up through generation 9 (3) using the molecular simulation facilities of **POLYGRAF** (from Biodesign, Inc., Pasadena, CA 91101). These investigations indicate a dramatic change in the overall structural properties for the β-alanine systems as these polymers grow past generation 4. The early generation dendrimers

Figure 2. Cascade growth pattern for the β-alanine type dendrimers from the core ammonia unit up to generation 4.

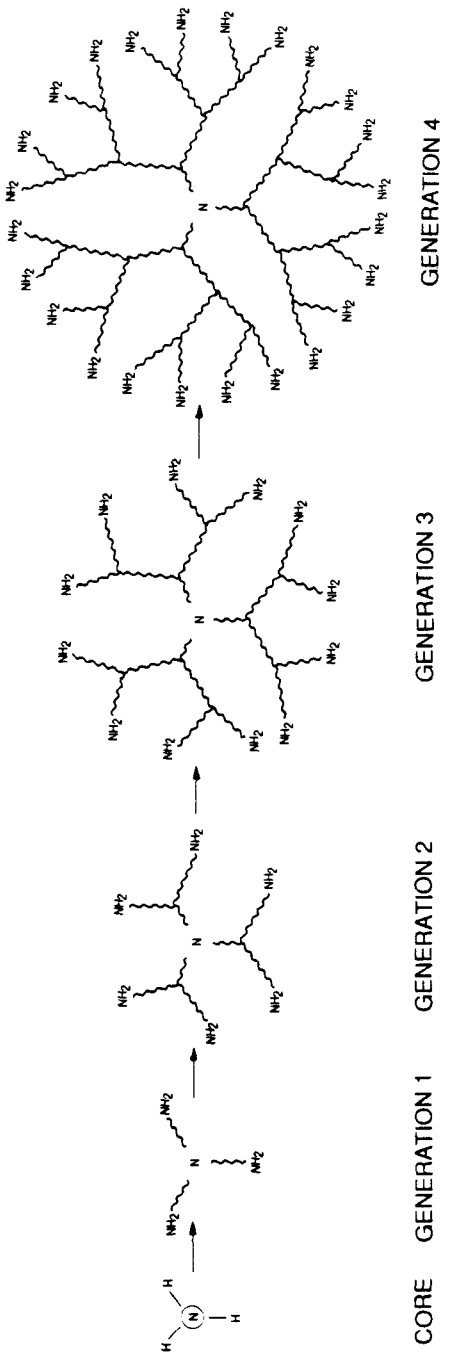
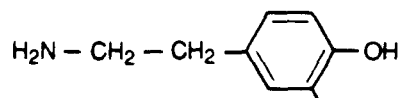


Figure 2.

(generations 1 through 3) possess very open and well-extended structures. These structures are hemispherical disks (with the core nitrogen responsible for the curvature) but there is no real inside to the polymer. This topology for the early generation systems is strikingly different from earlier proposals (4), but these hemispheric structures provide a simple interpretation of observed nuclear magnetic resonance relaxation times (Naylor, A. M.; Goddard III, W. A.; Keifer, G. E.; Tomalia, D. A. J. Am. Chem. Soc., in press 1988).

We find a distinct change in structural properties for those polymers above generation 5. The overall structure of the polymers is more sphere-like, with ample internal hollows connected by channels that run the length of the polymer assembly. These interior cavities should be suitable for sequestering guest molecules. Thus, in Figure 3, which limns a generation 6 dendrimer, we find an overall spherical structure containing a channel running through the interior of the polymer from one side to the other! To better illustrate the space available in the interior, we show in Figure 4 the solvent-accessible surface for the polymer (generation 6). [The solvent-accessible surface is constructed using the approach developed by Michael Connolly (University of California, San Diego) and is based on the definitions for molecular surfaces proffered by Frederic Richards (5). The molecular surfaces are composed of dots located in terms of a probe sphere (radius = 1.4 Å for water) rolling along the van der Waals spheres of the outer (accessible) atoms of the polymer.] Here we have used graphical slabbing to remove from view the front and rear portions of the image. The channels and cavities in this structure are representative of those found in the higher generation β-alanine dendrimers.

Dopamine – Sequestered Dendrimers. As an illustration of the dimensions of these channels and of how dendrimers might be used to sequester small molecules, we have used the molecular simulation capabilities of **POLYGRAF** to predict the optimum conformations for several dopamine **1** molecules



DOPAMINE

1

inside this starburst polymeric matrix. Dopamine was chosen because it is a good candidate for effective sequesterization in a β-alanine type dendrimer. Its size and shape are suitable for the cavities and channels found to exist in the higher generation β-alanine systems. From a chemical standpoint, dopamine possesses both the hydrogen bond donor and acceptor sites needed for favorable binding interactions to the polymer's carbonyl, amide, and amino substituents. We estimate that a generation 6 polymer should be capable of holding 15-20 molecules of dopamine. Figure 5 illustrates a dopamine/dendrimer complex with examples of the conformations adopted by the sequestered molecules and the polymeric material.

Figure 3. A representative structure for a generation 6 β-alanine dendrimer where generations 1 to 4 are colored red, generation 5 is shown in light blue, and generation 6 is yellow.

Figure 4. A view of a generation 6 dendrimer with its solvent-accessible surface displayed and clipped to reveal the nature of the internal cavities and channels found in the higher generation β-alanine dendrimers.

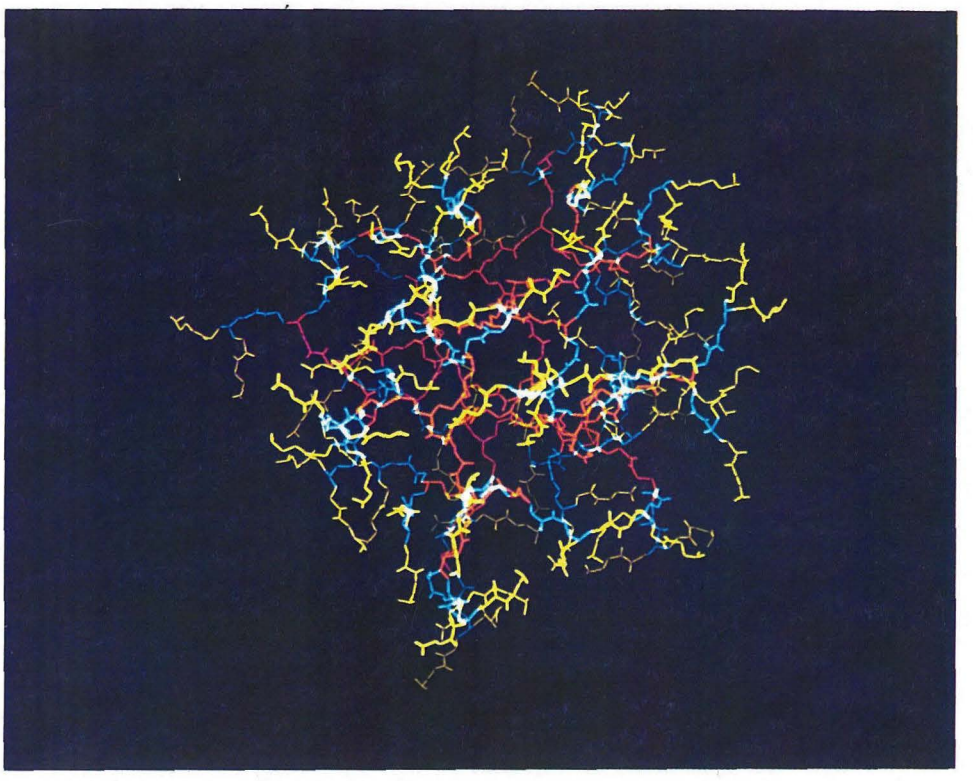


Figure 3.

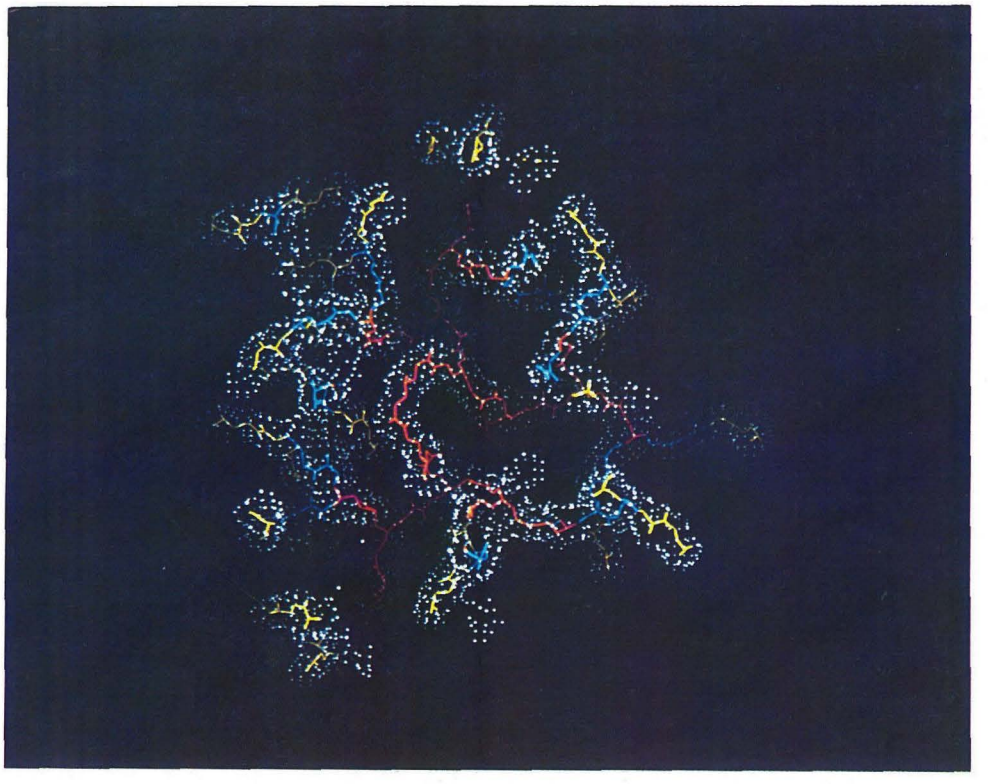


Figure 4.

Figure 5. A view of various dopamine molecules (magenta) bound in the inner regions of a β-alanine dendrimer (generations 1 through 5 in light blue; generation 6 in yellow).

Figure 7. A catechol-modified, dopamine-loaded β-alanine dendrimer where the dopamines are displayed in magenta, generations 1 through 5 in light blue, generation 6 in yellow, and the catechol cap in red.

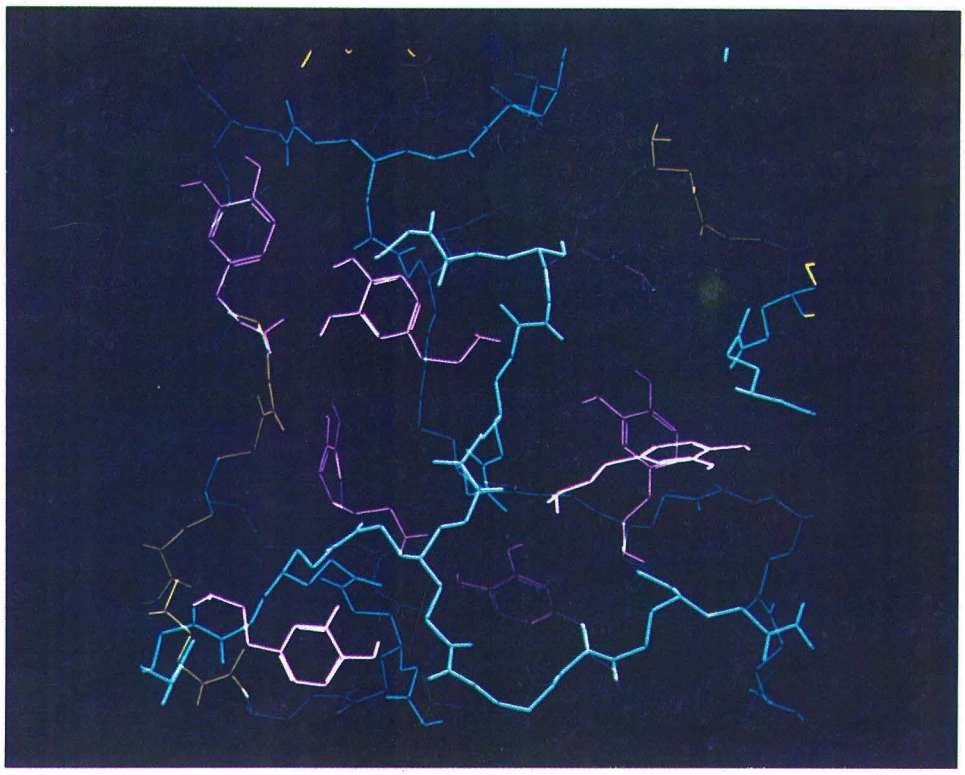


Figure 5.

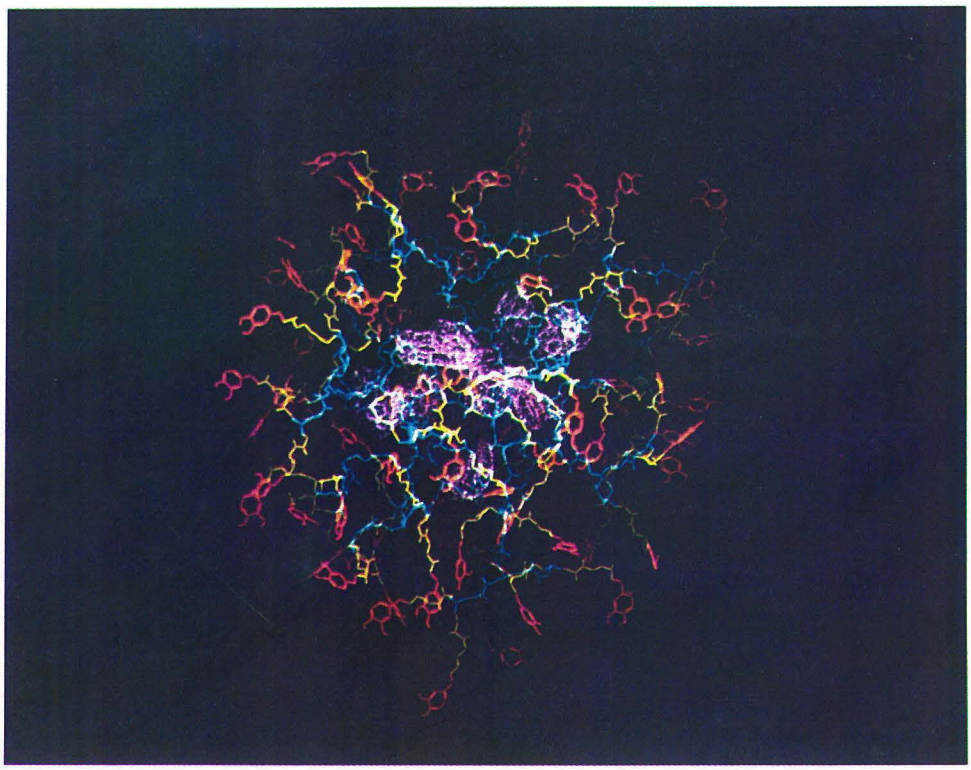
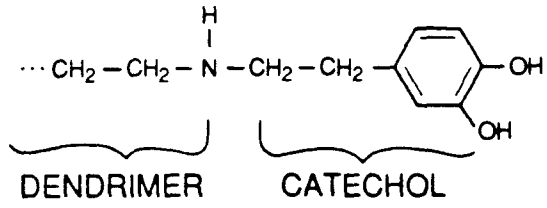


Figure 7.

Pharmacologically, the ability (i) to selectively deliver dopamine to peripheral kidney receptors without eliciting complications due to the presence of dopamine receptors in the central nervous system (CNS) and (ii) to maintain a supply of nonmetabolized dopamine would be advantageous in cardiovascular hypertension therapy (6). The additional issues presented by this type of application include the dimensions of the dendrimer encapsulated dopamine and the targetting of the unit to the appropriate dopamine receptor sites.

The dimensions of the higher generation β-alanine dendrimers (see Figure 6 with the minimum diameter plotted as a function of generation) are such that these dopamine/dendrimer complexes would not readily cross the CNS blood-brain barrier.

In order to ensure the delivery to kidney dopamine receptors, we suggest modifying the terminal amine sites of the polymer with catechol fragments, as shown in **2**



2

This leads to surface features resembling dopamine and may lead to selective binding at dopamine receptor sites (Figure 7). The polymer encapsulation matrix should protect the catecholamine from rapid metabolic inactivation.

These studies provide structural models that should be useful for analyzing the dopamine/dendrimer systems. The next step is to test the effectiveness of these modified materials for encapsulation of dopamine (and related materials) and to determine how effectively they are delivered to the kidney centers. As such experiments proceed, continuing simulation will be useful in providing a quantitative framework for understanding various results.

We expect that the judicious selection of a core unit, internal monomer subunits, and terminal monomer fragments will allow the design of dendrimers to complex with specific guest molecules and to deliver them to specific sites. It will, of course, be essential to develop synthetic schemes that allow for chemical control and fidelity.

Dihydrofolate Reductase : Roles of Precise Structure in Catalysis

In designing new catalysts, it is essential to understand how the character of the active site (including cofactor) controls catalytic activity since one will want to modify this active site or cofactor. As a prototype for such studies, we have carried out a series of molecular simulations on the Dihydrofolate Reductase (DHFR) system.

Figure 6. A plot of the β-alanine dendrimers "short" diameter (see text) as a function of generation. These diameters were calculated from our molecular simulations and based on the average of the smallest principal moment of inertia.

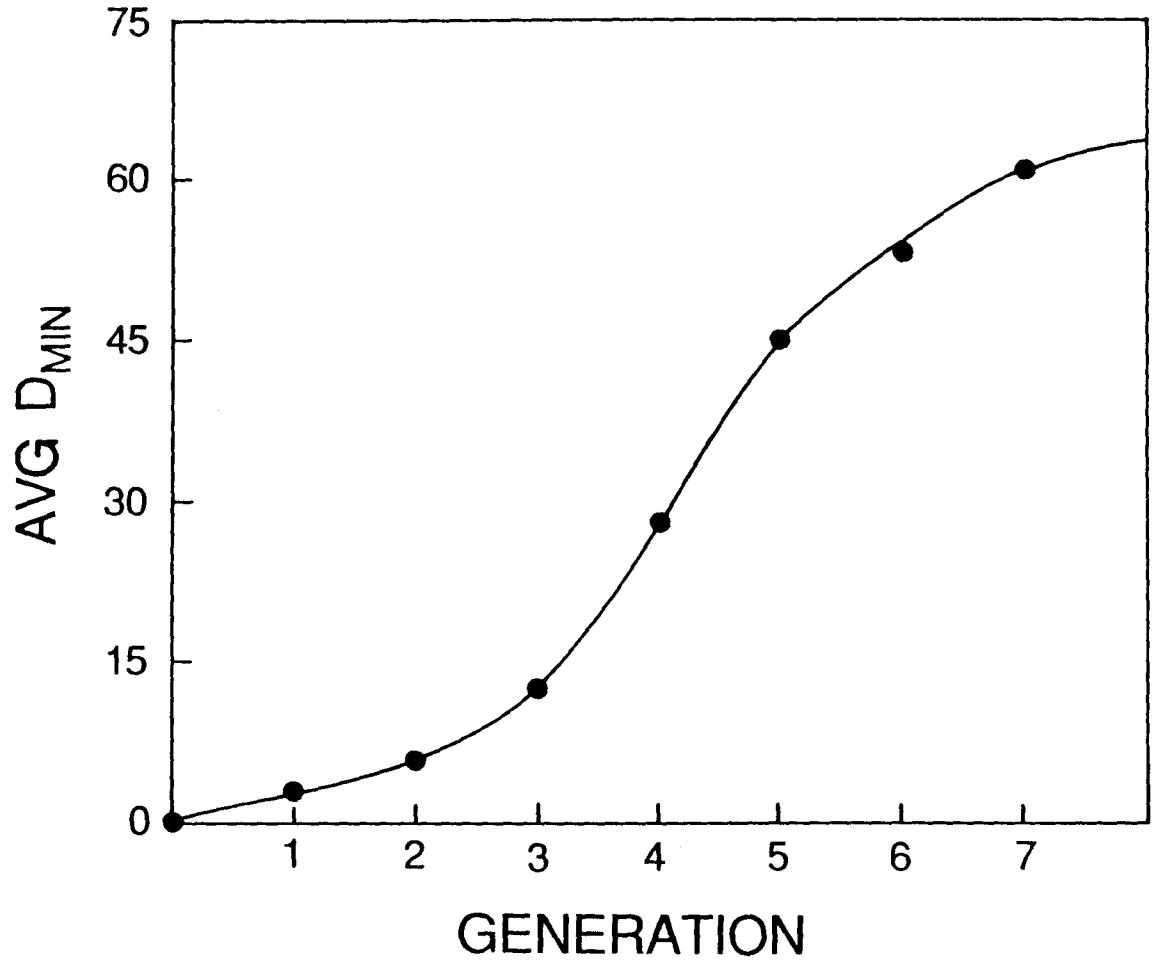
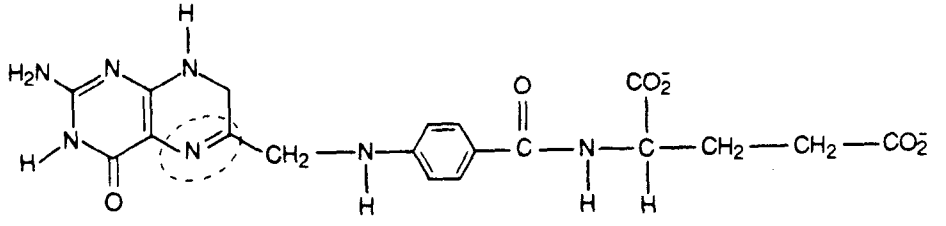


Figure 6.

Introduction and Rationale. DHFR is an ideal system to study for a number of reasons. The catalytic properties of DHFR are such that under normal physiologic conditions and with the NADPH cofactor bound, 7,8-dihydrofolate (DHF) is reduced to 5,6,7,8-tetrahydrofolate (THF) (7). Thus DHFR plays an important role in cell metabolism by maintaining a supply of THF. THF is used by the cell as both a cofactor and in substrate quantities in the synthesis of deoxythymidine. By inhibiting the production of THF, deoxythymidine synthesis is curtailed, nucleic acid replication comes to a halt, and cell proliferation ceases. It is this biochemical cascade which supplies the pharmacological and chemotherapeutic applications of inhibitors to DHFR.

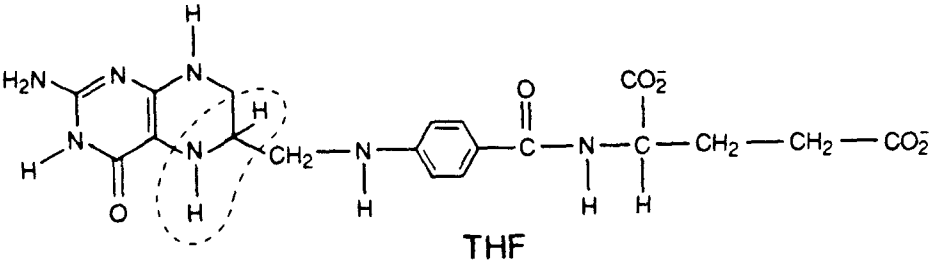
The abundant experimental data on DHFR provide both inspiration and detailed checks upon theoretical studies. The crystal structures of DHFR from two bacterial sources have been resolved, reported, and coordinates made available through the Brookhaven Data Base (8-10). The kinetic profiles of the enzyme under various conditions and, hence, mechanistic details have been determined (11,12). The recombinant DNA protocols for the successful study of site-directed mutants of DHFR have been reported (13-17).

DHF **3** is composed of (from left to right, below) a pterin ring, a bridging methylene unit, p-aminobenzoate, and a glutamate fragment.



DHF

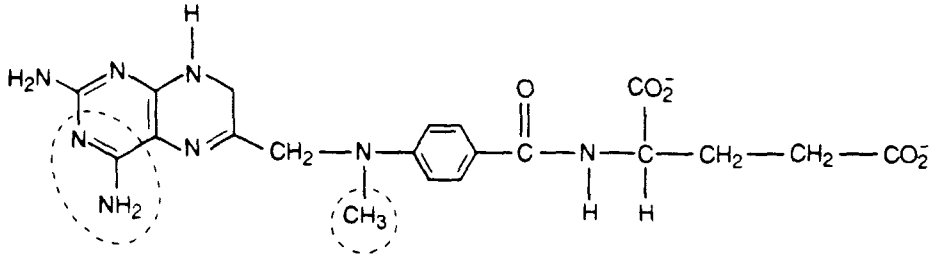
3



THF

4

It is the N5-C6 double bond (circled, above) of 7,8-DHF that is reduced to 5,6,7,8-THF (4) by DHFR. Methotrexate (MTX), an important chemotherapy and anti-bacterial agent, is chemically quite similar to DHF (differences are circled).



MTX

5

MTX **5** contains the same benzglutamate fragment as DHF (but with a methylated amine) and a pteridine ring with an amino substitution at position 4. Indeed, we find smilar binding sites for MTX and DHF on DHFR.

The crystal structure of the *E. coli* DHFR/MTX binary complex (9) reveals that the inhibitor binds in the active site in a kinked fashion (Figure 8). Asp-27 forms a salt-bridge to the bound pteridine ring while Phe-31 allows for the bend in the bound inhibitor. The remainder of the active site cavity surrounding the pterin ring is composed of amino acid residues that create a very hydrophobic environment.

Before embarking on designing improvements into the catalytic sequence, it is useful to know the detailed kinetics of the system. A profile of the steady-state kinetics of the *E. coli* form of DHFR was recently reported by Fierke, Johnson, and Benkovic (12). This study used stopped flow fluorescence and absorbance spectroscopies to measure the rates of ligand association and dissociation and the binding constants of key intermediates in the catalytic sequence, as diagrammed in Figure 9. This study revealed that the catalytic step, i.e., the N5-C6 double bond reduction, proceeds at a steady-state rate of 950 sec^{-1}, whereas the rate of THF product release was only 12.5 sec^{-1}, making it the rate-limiting step. Hence, any design modifications to increase the turnover of this system should focus first upon increasing the product off-rate.

A Synthetic Mutant : Phe − 31 → Tyr − 31. We now turn to using molecular simulation methods to examine some areas of the biocatalytic properties of DHFR that are not easily investigated experimentally. The general methods used for our studies incorporate computer graphics, molecular force fields, energy minimization, and molecular dynamics. **BIOGRAF** (from Biodesign, Inc., Pasadena, CA 91101) was the molecular simulation package used for these investigations. The AMBER force field was used for the calculations (18).

As an illustration of the use of simulations, we consider the effect on rate of a site mutation of DHFR. Benkovic *et al.* used recombinant-DNA techniques to synthesize the protein where Phe-31 was transformed into Tyr-31 (14). Preliminary experimental studies indicated a change in the pK_a of Asp-27 in this mutant form of the enzyme. Molecular simulations were run on both the wild-type enzyme and the Tyr-31 mutant to assess the structural changes produced by such a modification.

Figure 8. The active site of *E. coli* DHFR with MTX and NADPH bound.

Figure 10. A comparison of MTX bound to wild-type DHFR (red MTX and Phe-31; blue active site residues) and Tyr-31 mutant protein (yellow MTX, Tyr-31, and active site residues).

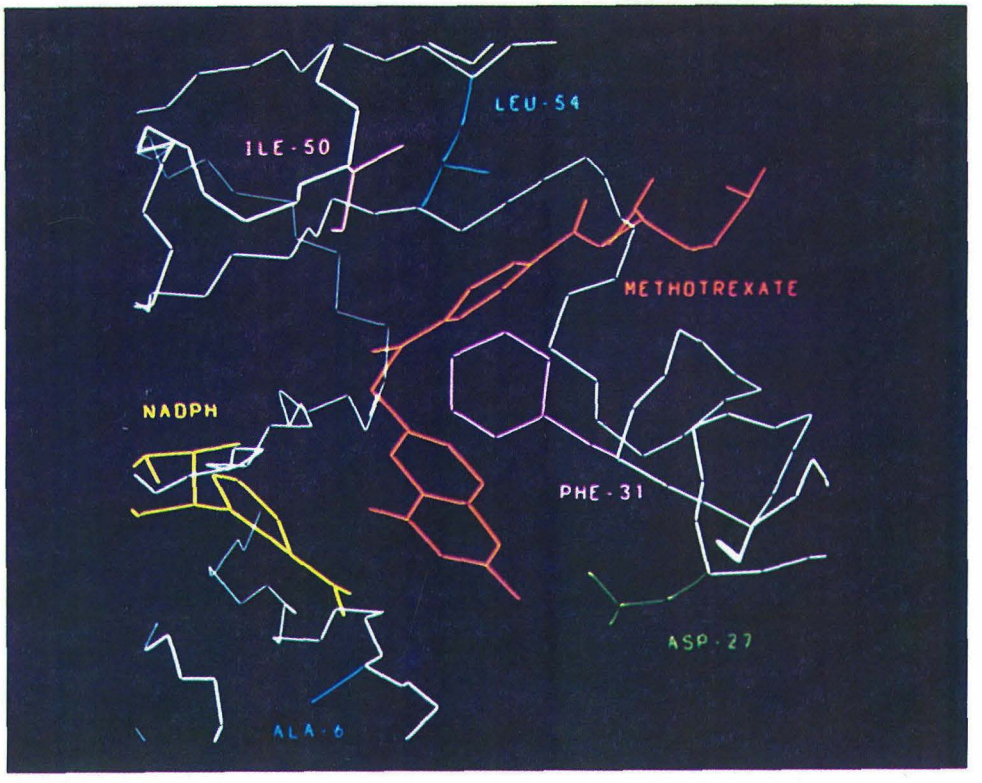


Figure 8.

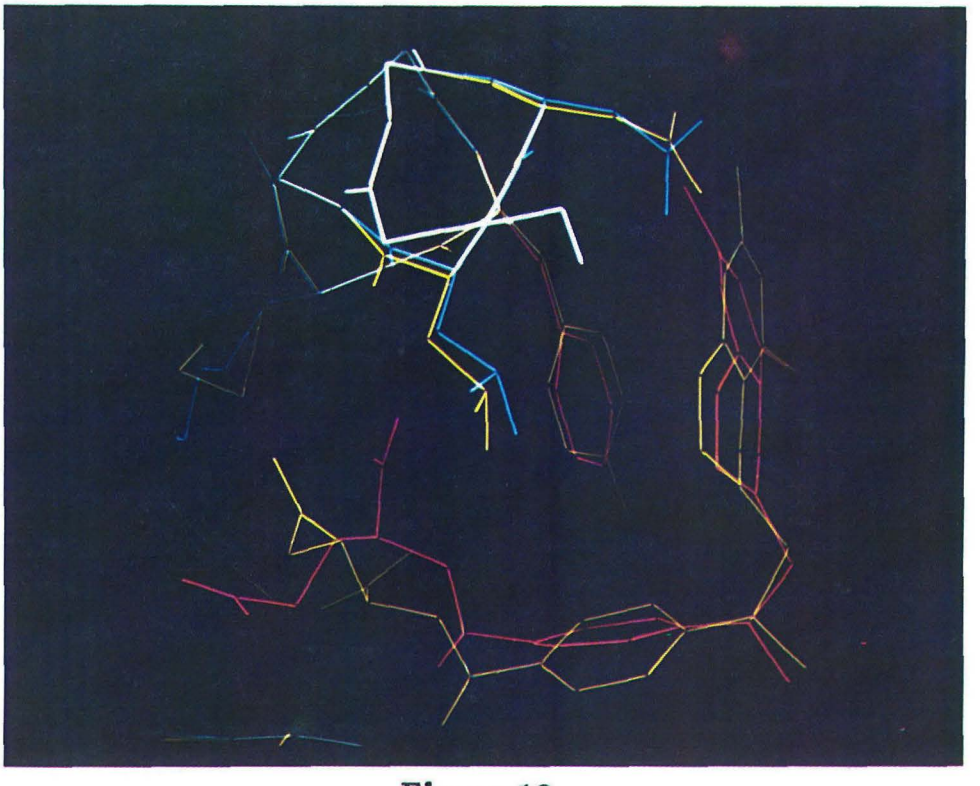


Figure 10.

Figure 9. Steady-state kinetic cycle for the reduction of 7,8-DHF to 5,6,7,8-THF via DHFR.

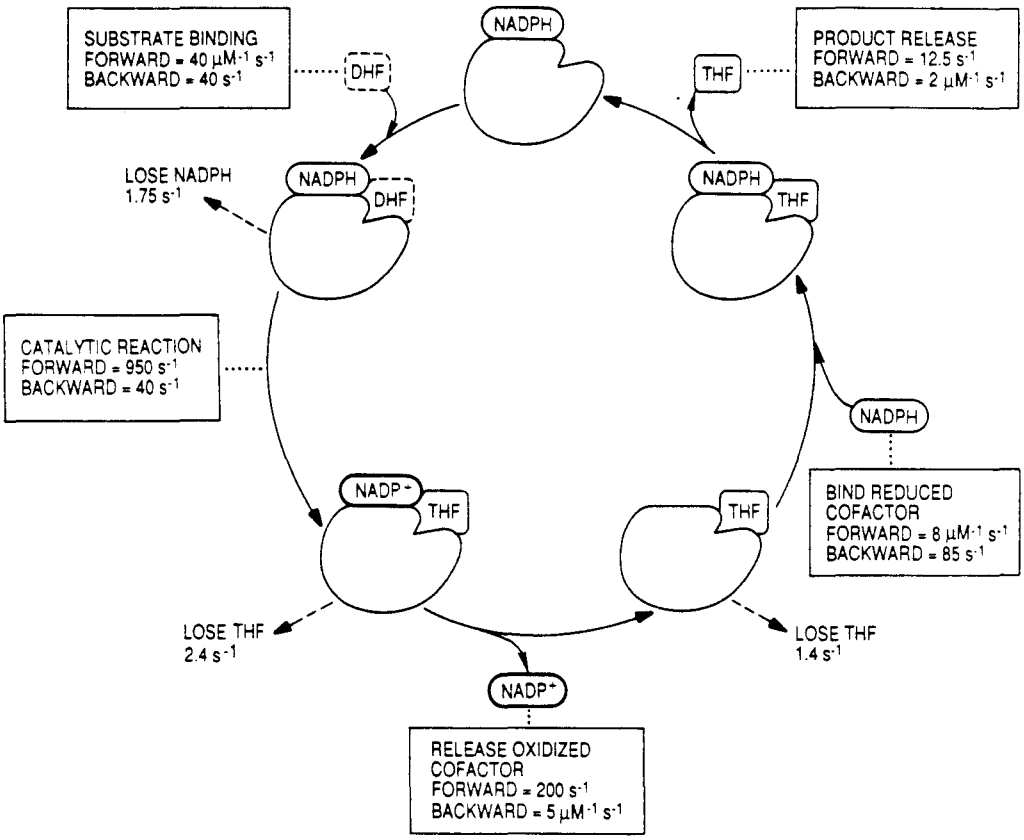

SUBSTRATE BINDING
FORWARD = 40 μM⁻¹ s⁻¹
BACKWARD = 40 s⁻¹
DHF
NADPH
PRODUCT RELEASE
FORWARD = 12.5 s⁻¹
BACKWARD = 2 μM⁻¹ s⁻¹
THF
LOSE NADPH
1.75 s⁻¹
NADPH
DHF
NADPH
THF
CATALYTIC REACTION
FORWARD = 950 s⁻¹
BACKWARD = 40 s⁻¹
NADPH
BIND REDUCED
COFACTOR
FORWARD = 8 μM⁻¹ s⁻¹
BACKWARD = 85 s⁻¹
NADP⁻
THF
THF
LOSE THF
2.4 s⁻¹
LOSE THF
1.4 s⁻¹
NADP⁺
RELEASE OXIDIZED
COFACTOR
FORWARD = 200 s⁻¹
BACKWARD = 5 μM⁻¹ s⁻¹

Figure 9.

The starting structure for our wild-type protein complex was based upon the crystallographic data, where the *L. casei* NADPH coordinates were used to model the cofactor into the *E. coli* system. To create the Tyr-31 mutant, the replace facility of **BIOGRAF** was used. For both series of simulations, a 10 Å solvent shell (containing experimental plus supplemental waters) surrounded the protein complexes. Counter ions were positioned near charged amino acid side chains to achieve electroneutral systems. The structures were then equilibrated via a sequence of molecular dynamics and energy minimization calculations.

Our calculations show that the pteridine ring is relatively undisturbed by the introduction of the Tyr hydroxyl group, yet the benz-glutamate portion of the bound substrate (or inhibitor) is markedly displaced (see Figure 10). We find that the mutant -OH group produces a 'bump' on the surface of the hydrophobic pocket near the benzyl group that disrupts the binding of the benz-glutamate fragment. This suggests that the product off-rate from the DHFR/NADPH/THF complex should be affected markedly by this substitution. Kinetic studies have verified that the product off-rate for the Tyr-31 form of *E. coli* DHFR is 50 times faster than that of wild-type protein (14)!

Natural Mutants : *E. coli* and *L. casei*. Consider now the question of how different one can make the active site *without* affecting the chemistry. The overall amino acid sequence homology between the *E. coli* and *L. casei* forms of DHFR is only 27%. However, as shown in Figure 11, a Gibbs free energy comparison of the steady-state data for the *E. coli* and the *L. casei* proteins shows that the reaction kinetics for these two forms of the enzyme are remarkably similar (19). In fact (using the ternary DHFR/NADPH/DHF complex as the energy reference), the energy differences between the two are, at every step, within 1 kcal/mol! Using the x-ray crystal structures and reorienting them so that the MTX of both complexes are superimposed (by docking one structure with respect to the other), we see that the conformation of MTX bound to the *E. coli* DHFR is very similar to that of MTX in the *L. casei* ternary crystal complex. The orientations of both the pteridine and the benzyl rings are identical for the two complexes and the backbone atoms of the glutamate substituent also align quite nicely. The discrepancy in these two forms of bound MTX occurs at the C-δ carboxylate group of the glutamate fragment, a position far removed from the catalytic site. A glance at the two different enzymes reveals that this difference in conformation is due to the positions of positively charged amino acid side chains on opposite sides of the opening for the active site (Arg-52 for the *E. coli* form and His-28 for the *L. casei* complex).

The results of docking the *L. casei* structure to the *E. coli* show that the overall folds for these two proteins are quite similar. Regions of secondary structural elements, i.e., α-helices and β-sheets, are conserved, while insertions and deletions in the amino acid sequences occur in loop regions located about the exterior regions of the enzyme. The structural domains of the DHF and NADPH binding sites are maintained. In the comparison of the DHF substrate binding pockets, one finds that nature, in some instances, substitutes like for like by using

Figure 11. A Gibbs free energy plot of the kinetic profiles for *E. coli* (solid) and *L. casei* (dotted) DHFR (19).

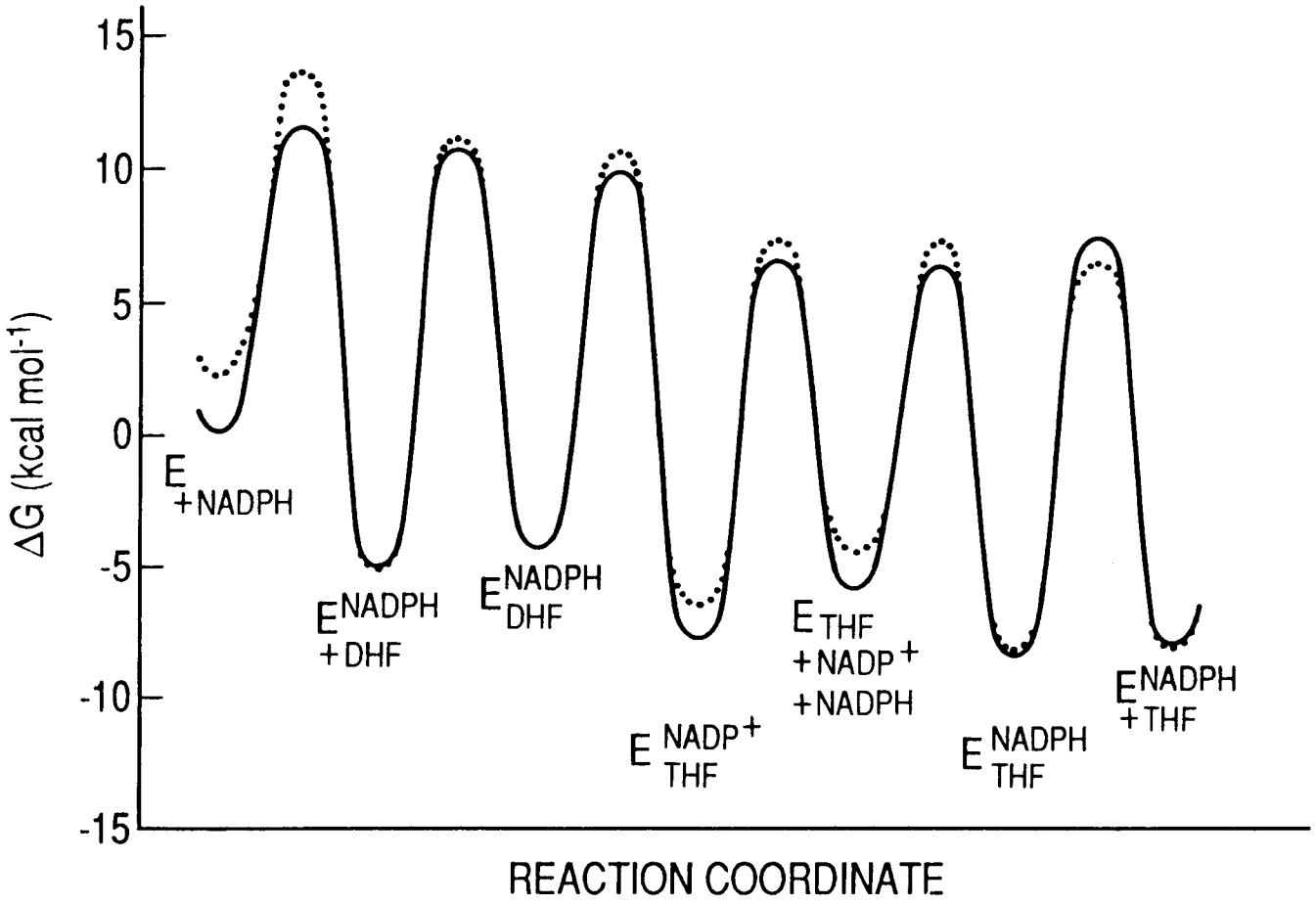

15
10
5
0
-5
-10
-15
ΔG (kcal mol-1)
E +NADPH
E NADPH +DHF
E NADPH DHF
E NADP+ THF
E THF +NADP+ +NADPH
E NADPH THF
E NADPH +THF
REACTION COORDINATE

Figure 11.

(1) different aromatic side chains [as Trp-30 (*E. coli* ≡ *Ec*) → Tyr-29 (*L.casei* ≡ *Lc*), Tyr-100 (*Ec*) → Phe-103 (*Lc*), and Phe-153 (*Ec*) → Tyr-155 (*Lc*) indicate] or

(2) different aliphatic side chains [as Ile-5 (*Ec*) → Leu-4 (*Lc*), Leu-36 (*Ec*) → Val-35 (*Lc*), and Leu-112 (*Ec*) → Val-115 (*Lc*) demonstrate].

But there also are examples where aliphatic is substituted for aromatic (and vice versa) as with Ile-50 (*Ec*) → Phe-49 (*Lc*) and Tyr-111 (*Ec*) → Leu-114 (*Lc*). Similar substitution patterns are found in the NADPH cofactor binding site, where, in addition to those types described above, charged residues are maintained [i.e., His-45 (*Ec*) → Arg-44 (*Lc*) and Lys-76 (*Ec*) → His-77 (*Lc*)].

Using these superimposed structures with the MTX (from both crystal structures) and the NADPH (from the *L. casei* complex) as guides, the active site regions are identified for both enzymes. Here we define the active site to be those amino acids containing any atom within 7 Å of either of the bound substrates. Analyses of the residues in the active site region determine that the amino acid sequence homology in this portion of the protein is 35% (see Table I).

The two docked active site regions (with MTX bound) are shown in Figure 12. Examination of the docked active sites reveals that there are three general types of amino acid residues involved in the construction of the active site surface:

(1) amino acid residues conserved between the two bacterial forms of DHFR, e.g., Phe-31 (*Ec*) and Phe-30 (*Lc*);

(2) amino acid residues that differ between the two sequences and contribute chemically different side chains to the construction of the active site, e.g., Ile-50 (*Ec*) and Phe-49 (*Lc*); and

(3) amino acid residues that differ between the two sequences but contribute only chemically homologous main chain atoms to the construction of the active site, e.g., Ala-6 (*Ec*) and Trp-5 (*Lc*).

Thus, by modifying the definition of homology to also include those residues in the active site supplying only backbone atoms, one obtains a generalized 'chemical homology' of the active site. Indeed, the 'chemical homology' between the *E. coli* and the *L. casei* substrate binding pockets becomes 60% (19)!

A comparison of the solvent accessible surfaces [as described above; (5)] for these segments of the protein further substantiates these remarkable structural similarities. The surface areas mapped out for these regions agree within 93%. The solvent accessible sufaces for the MTX pteridine and the NADPH nicotinamide rings are essentially interchangeable for the two different sequences, as illustrated in Figure 13.

Hence, we see that nature creates essentially identical reactive surfaces for these bacterial DHFR systems from different amino acid building blocks (19) (Benkovic, S. J.; Adams, J. A.; Fierke, C. A.; Naylor, A. M. Pteridines, in press 1988). The components necessary for DHFR's ability to catalytically reduce the N5-C6 double bond of 7,8-DHF to 5,6,7,8-THF are conserved. The carboxylate group that supplies a proton to the pterin ring is present in both the *E. coli* and

Table I. Comparison[a] of DHFR Active Site Residues from *E. coli* and *L. casei*.

Residue[b]	4	5	6	7	8	9	10	...	13	14	15	16	17	18
E. coli	L	i	A	**A**	L	A	V	...	V	i	**G**	M	e	n
L. casei	F	l	W	**A**	Q	N	R	...	L	l	**G**	K	d	g
	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	a	m	**P**	**W**	n	**L**	**P**	A	**D**	**L**	a	w	**F**	k
	h	l	**P**	**W**	h	**L**	**P**	D	**D**	**L**	h	y	**F**	r
	33	34	35	36	...	42	43	44	45	46	47	48	49	50
	r	n	**T**	l	...	m	**G**	**R**	h	**T**	W	**E**	**S**	i
	a	q	**T**	v	...	v	**G**	**R**	r	**T**	Y	**E**	**S**	f
	51		52	53	54	55	56	57	...	61	62	63	64	65
	g	...	**R**	**P**	**L**	**P**	G	**R**	...	i	**L**	s	s	**Q**
	p	k	**R**	**P**	**L**	**P**	E	**R**	...	v	**L**	t	h	**Q**
	66	67	68	...	74	75	76	77	78	79	...	93	94	95
	p	G	t	...	w	**V**	K	S	**V**	D	...	V	I	**G**
	e	D	y	...	v	**V**	H	D	**V**	A	...	I	A	**G**
	96	97	98	99	100	101	102	103	...	111	112	113	114	115
	G	G	r	v	y	**E**	q	**F**	...	y	l	**T**	H	i
	G	A	q	i	f	**T**	a	**F**	...	l	v	**T**	R	l
	122	123	124	125	...	153								
	D	**T**	**H**	f	...	f								
	D	**T**	K	m	...	y								

[a] Bold upper case indicates homologous, lower case indicates nonhomologous, and nonbold upper case indicates backbone chemical homology, as described in the text.

[b] Numbered for the *E. coli* form.

Figure 12. A comparison of the docked *E. coli* (red MTX and enzyme) and *L. casei* (yellow MTX and NADPH; blue enzyme) active sites. The conserved Phe's are shown to the right of the pteridine rings. The substituted side chains [Ile-50 (*Ec*) and Phe-49 (*Lc*)] appear in the 12 o'clock position above the benzyl ring. The chemically homologous main chain contributors [Ala-6 (*Ec*) and Trp-5 (*Lc*)] are shown in the 7 o'clock position.

Figure 13. Dotted surface comparison showing the congruence of the *E. coli* (red) and *L. casei* (blue) active sites.

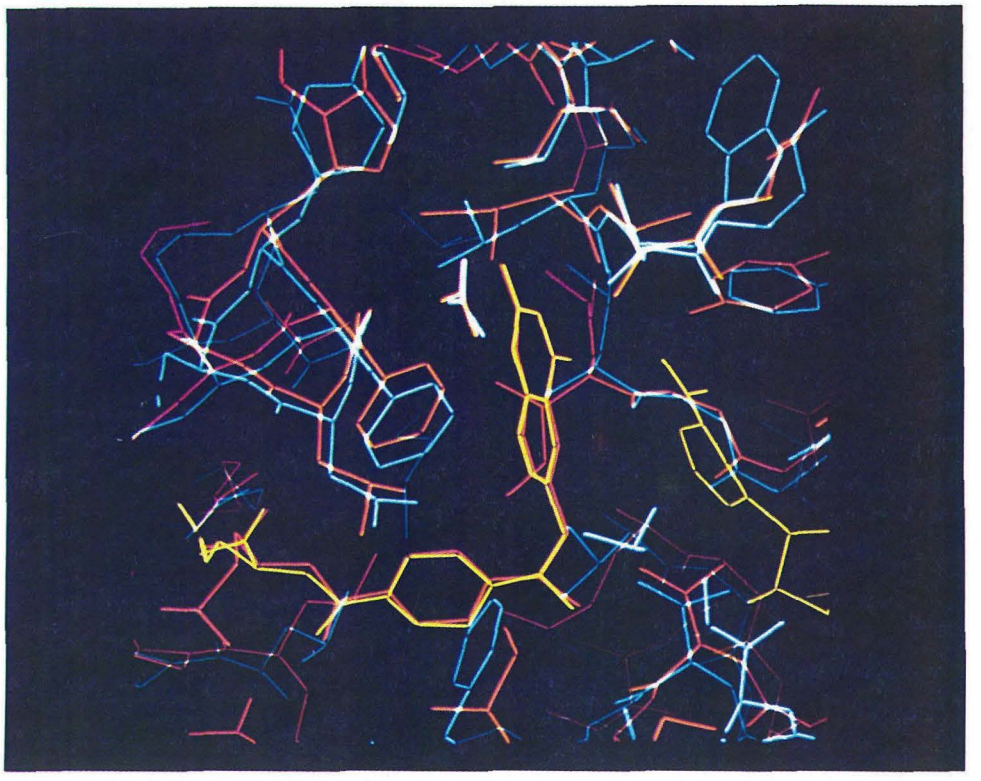


Figure 12.

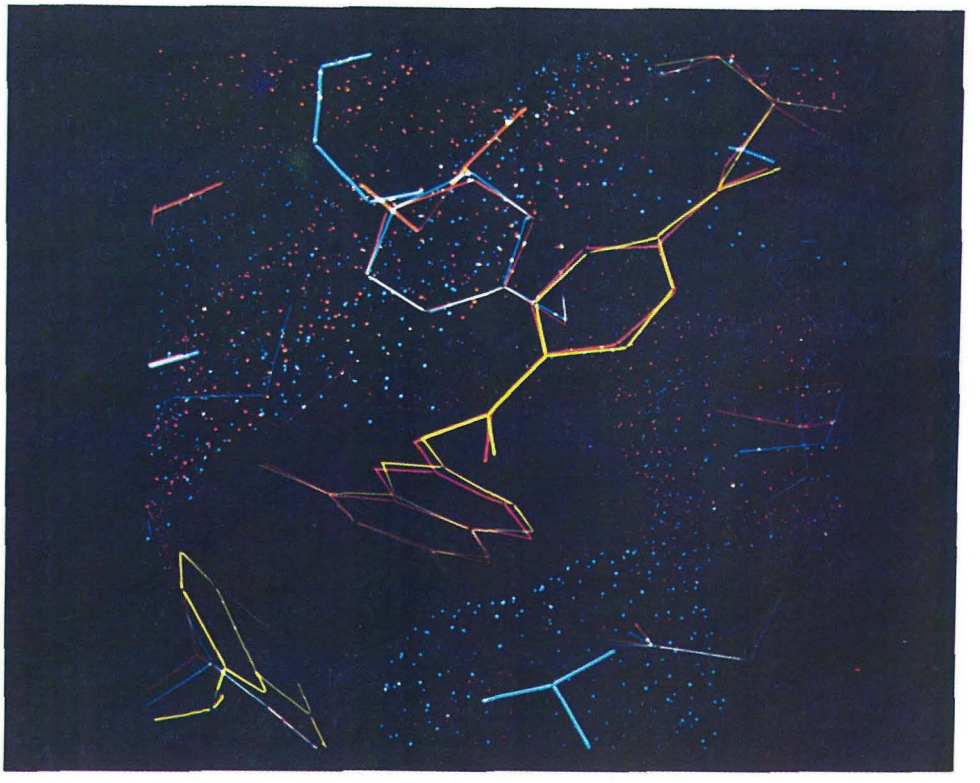


Figure 13.

the *L. casei* sites as an Asp side chain. The hydrophobic pocket that surrounds the reactive pteridine moeity and prevents the aqueous environment from disturbing the catalytic process is formed by varied amino acids but maps out the same surface area and hydrophobic character. The hydride donating NADPH nicotinamide ring is positioned adjacent to the reactive center even though the NADPH binding sites are composed of different amino acid sequences.

Conformation of Bound Inhibitor. Another area under investigation focuses on the conformation(s) of the substrate and an inhibitor (MTX) bound to DHFR. The form of the DHFR/MTX complex is known from the crystallographic studies of Kraut *et al.* (8-10). However, the orientation of the bound pterin ring in the reactive DHFR/DHF is known to differ dramatically from the MTX crystal structure (20). Basically, these differences arise because there are two possible orientations of the pterin ring in the active site; one is flipped by 180° with respect to the other. Isotope labelling experiments on THF show that the reactive DHF must be bound in the conformation flipped from that observed by x-ray for MTX. In order to understand these differences, we ran simulation studies on altered forms of bound MTX and DHF to investigate the structural and energetic properties of these systems.

The starting structures for our binary DHFR/MTX calculations were based upon the crystallographic data. For the experimental form of bound MTX we used the coordinates as they were reported. For the pteridine-flipped form of bound MTX, termed MTX-flipped, we rotated about the C6-C9 and C9-N10 bonds of MTX to produce a starting conformation with the pteridine ring positioned in the active site but with its face flipped 180° (Figure 14). To be consistent with experiment, the bound MTX inhibitor was protonated at N1 (21). For both systems we used the experimentally determined solvent plus additional water to create a 10 Å solvent shell around the enzyme-inhibitor complex. Electroneutral systems were achieved by positioning the appropriately charged counter ions near to cationic and anionic side chains. The entire solvated protein and inhibitor systems were then equilibrated with a sequence of molecular dynamics and energy minimization calculations (using **BIOGRAF** and the AMBER force field).

These simulations indicate that the active site of DHFR can accommodate both forms of the bound MTX. After equilibration of both the experimental and the flipped forms, the carboxylate group of Asp-27 is found to be involved in a salt bridge with the pteridine ring of the inhibitor. The hydrophobic pocket surrounding the pterin ring has adjusted adequately to either conformation of the bound MTX. The energetic profiles indicate that the experimentally observed form of MTX bound to DHFR is ~3 kcal/mol lower in energy than the flipped form of the bound inhibitor.

Conformation of Bound Substrate. Now the question of the conformation of bound DHF substrate was addressed. Simulations were run on DHFR/DHF/NADPH ternary complexes to investigate both the structural and energetic requirements associated with the reactive and nonreactive forms of DHF (see Figure 14). As above, the crystallographic information available for

Figure 14. The orientations of the MTX and DHF pteridine rings in the active site of DHFR.

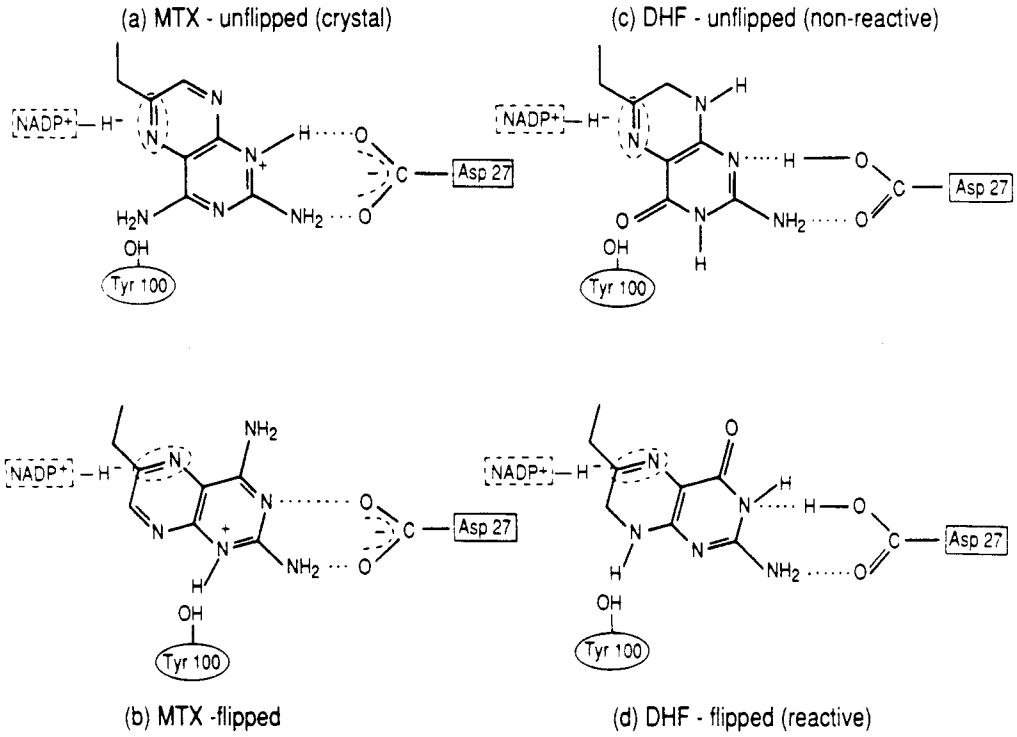

(a) MTX - unflipped (crystal)
(c) DHF - unflipped (non-reactive)
NADP+
H-
Asp 27
Tyr 100
(b) MTX -flipped
(d) DHF - flipped (reactive)

Figure 14.

the *E. coli* binary and *L. casei* ternary complexes (with MTX) was used to generate starting structures for the molecular simulations. The coordinates for the *E. coli* bound NADPH were modeled based upon those from the *L. casei* ternary complex. The DHF units were positioned based upon the coordinates of the MTX. For these simulations, the carboxylate group of Asp-27 was protonated while the pterin ring of the substrate was not protonated at the N1 position. Again, solvent and counterions were included and all atoms of the structures were relaxed and equilibrated during the calculations.

The simulations indicate that there are two favorable conformations of bound DHF in the active site of DHFR, each analogous to the favorable conformations of MTX. The amino acid side chains that create the hydrophobic portion of the active site reposition themselves to allow for either form of the pteridine ring. In the reactive orientation (termed DHF-flipped), the Asp-27 COOH prongs the C4 oxygen and the N3 sites. In the nonreactive conformation (termed DHF-unflipped), the carboxylate group straddles the N1 and the N8 positions. The energetic results of the simulations find *the unflipped nonreactive form* (analogous to the x-ray form in MTX) of the pteridine ring *lower in energy than the reactive form* by ~4 kcal/mol! This result seems to contradict experiment since no radio-labeled product resulting from DHF reduction accommodates this orientation of bound substrate.

Closer examination of the active site regions reveals that the reducible C-N bond of the pterin ring differs substantially with respect to the positions of the NADPH nicotinamide rings. In the low energy nonreactive state, the NADPH C'4 site is 4.1 Å from the carbon of the pteridine N5-C6 double bond, while in the higher energy reactive form, this interatomic distance is reduced to only 3.2 Å! The shorter distance for hydride transfer during catalysis favors the reactive conformation even though the energetics slightly favor the nonreactive form of bound DHF. This information clearly suggests a kinetic mechanism for the control of DHF reduction to THF via DHFR. Figure 15 depicts the relationship between the NADPH nicotinamide ring and the reactive DHF substrate.

Wu and Houk have used ab initio calculations to propose a mechanism for hydride transfer to the methyl iminium cation (22). Their results indicate that H^- transfer proceeds via a syn transition state structure with a C-H-C angle of 150-160° and a C-C interatomic distance of 2.6 Å (see Figure 16). What we find for the reactive form of bound DHF and NADPH in an equilibrated ground state is a pseudo-syn orientation of the reacting ring systems, a C-H-C angle of 155°, and an equilibrium distance of 3.2 Å between the two carbon centers (19). In contrast to the nonreactive conformer of DHF, simulations suggest that the equilibrium distance between the two reacting carbons is 4.1 Å. The key difference in character is the orientation of the pterin and nicotinamide rings. In the productive conformation, the two-ring systems adopt positions that allow for H^- transfer from the nicotinamide center to C6, while this relationship is not accommodated by the nonreactive bound conformation of DHF.

In order to address the reaction rates of these species, we could monitor the C-C distance as a function of time. Taking the probability distribution of C-C

Figure 15. The catalytic pocket of DHFR showing the relationship between the NADPH nicotinamide ring (left) and the reactive DHF pteridine ring (right).

Figure 15.

Figure 16. Proposed model for hydride transfer from NADPH to DHF (19). θ is the angle between the C4′ of $NADP^+$, the H^- being transferred, and C6 of DHF.

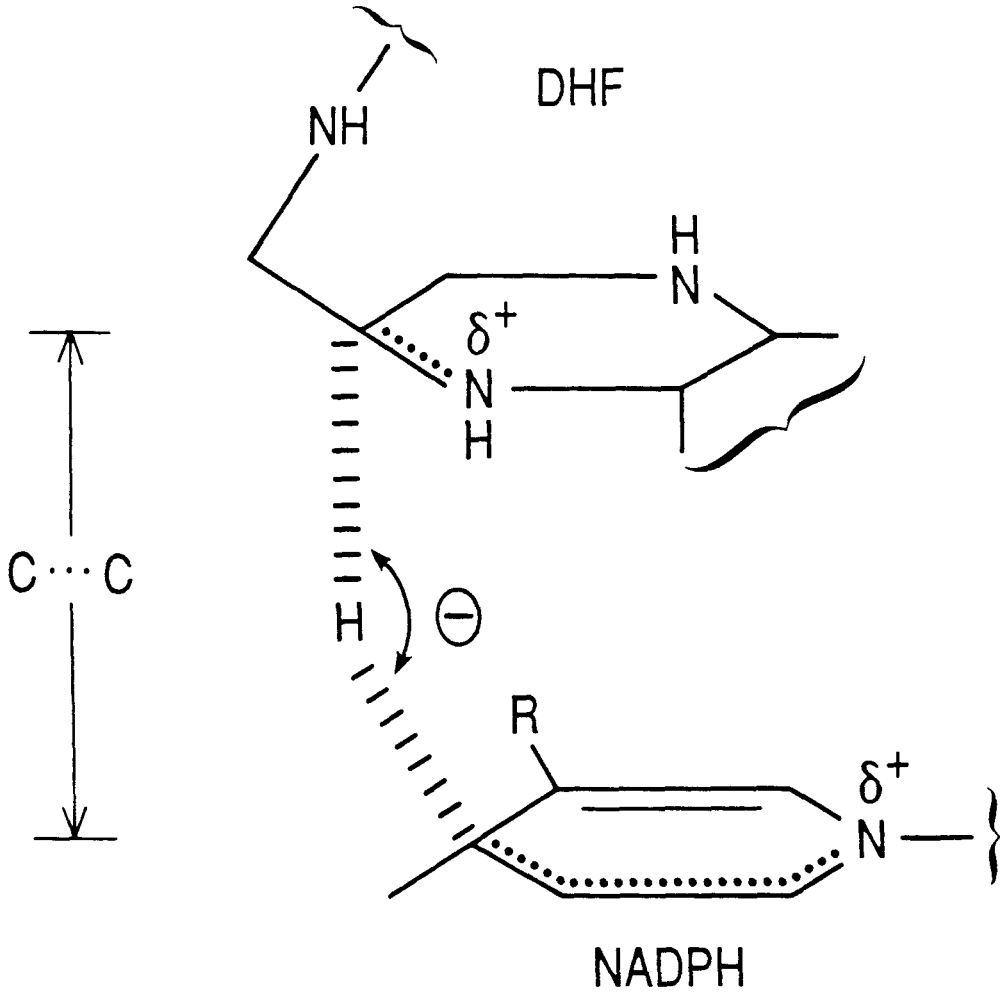


Figure 16.

with the reaction rate as a function of C-C distance, one should be able to predict relative reaction rates. However, qualitatively the large differences in equilibrium position certainly suggest a significant difference in rate. A test of this prediction would be to use NMR NOE experiments to examine conformation. We predict that over 90% of the bound complexes are the *non*reactive forms.

Summary

We have examined several systems chosen to illustrate the current role of theory and simulation in biomimetics and biocatalysis. It should be clear that the theory is not done in a vacuum (so to speak) but rather that the theory becomes interesting only for systems amenable to experimental analysis. However, the examples illustrate how the theory can provide new insights and deeper understanding of the experiments. As experience with such simulations accumulates and as predictions are made on more and more complex systems amenable to experiment, it will become increasingly feasible to use the theory on unknown systems. As the predictions on such unknown systems are tested with experiment and as the reliability of the predictions increases, these techniques will become true design tools for development of new biological systems.

Acknowledgments

This work was funded by a grant from the Department of Energy, Energy Conversion and Utilization Technologies. The DOE-ECUT program funded this work with the hope that development of simulation techniques would eventually have an impact on development of new, industrially useful processes for biotechnology and biocatalysis. We especially wish to acknowledge the foresight of Drs. Jim Eberhardt, Minoo Dastoor, and Jovan Moacanin in encouraging these efforts. The equipment used was also funded by the ONR/DARPA (Contract No. N00014-86-K-0735) and by a grant from the Division of Materials Research, Materials Research Groups, of the National Science Foundation (Grant No. DMR84-21119). We thank Professor Stephen J. Benkovic of the Pennsylvania State University and Dr. Donald A. Tomalia of The Dow Chemical Company for numerous helpful discussions and for materials in advance of publication.

Literature Cited

1. Cytochrome P – 450 : Structure, Mechanism, and Biochemistry; Ortiz de Montecellano, P. R., Ed.; Plenum Press: New York, 1986.
2. Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, M.; Kallas, G.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. Macromolecules 1986, 19, 2466-2468.
3. Naylor, A. M.; Goddard III, W. A. Polymer Preprints 1988, 29, 215-216.
4. Tomalia, D. A.; Hall, M.; Hedstrand, D. M. J. Am. Chem. Soc. 1987, 109, 1601-1603.
5. Richards, F. Ann. Rev. Biophys. Bioeng. 1977, 6, 151-176.
6. Kebabian, J. W.; Agui, T.; van Oene, J. C.; Shigematsu, K.; Saavedra, J. M. Trends in Pharmacological Sciences 1986, 96-99.
7. Gready, J. E. Adv. Pharmacol. Chemother. 1980, 17, 37-102.

8. Matthews, D. A.; Alden, R. A.; Bolin, J. T.; Filman, D. J.; Freer, S. T.; Hamlin, R.; Hol, W. T. J.; Kisliuk, R. L.; Pastore, E. J.; Plante, L. T.; Xuong, N; Kraut, J. J.Biol.Chem. 1978, 253, 6946-6954.
9. Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. J. Biol. Chem. 1982, 257, 13650-13662.
10. Filman, D. J.; Bolin, J. T.; Matthews, D. A.; Kraut, J. J. Biol. Chem. 1982, 257, 13663-13672.
11. Stone, S. R.; Morrison, J. F. Biochemistry 1984, 23, 2753-2758.
12. Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. Biochemistry 1987, 26, 4085-4092.
13. Villafranca, J. E.; Howell, E. E.; Voet, D. H.; Strobel, M. S.; Ogden, R. C.; Abelson, J. N.; Kraut, J. Science 1983, 222, 782-788.
14. Chen, J.-T.; Mayer, R. J.; Fierke, C. A.; Benkovic, S. J. J. Cellular Biochem. 1985, 29, 73-82.
15. Howell, E. E.; Villafranca, J. E.; Warren, M. S.; Oatley, S. J.; Kraut, J. Science 1986, 231, 1123-1128.
16. Taira, K.; Chen, J.-T.; Fierke, C. A.; Benkovic, S. J. Bull. Chem. Soc. Jpn. 1987, 60, 3025-3030.
17. Mayer, R. J.; Chen, J.-T.; Taira, K.; Fierke, C. A.; Benkovic, S. J. Proc. Natl. Acad. Sci. U.S.A. 1986, 83, 7718-7720.
18. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, Jr., S.; Weiner, P. J. Am. Chem. Soc. 1984, 106, 765-784.
19. Benkovic, S. J.; Fierke, C. A.; Naylor, A. M. Science 1988, 239, 1105-1110.
20. Charlton, P. A.; Young, D. W.; Birdsall, B.; Feeney, J.; Roberts, G. C. K. J. Chem. Soc., Chem. Commun. 1979, 922-924.
21. Cocco, L.; Temple, Jr., C.; Montgomery, J. A.; London, R. E.; Blakley, R. L. Biochem. Biophys. Res. Commun. 1981, 100, 413-419.
22. Wu, Y.-D.; Houk, K.N. J. Am. Chem. Soc. 1987, 109, 2226-2227.

Section II

Dendritic Macromolecules:

Molecular Simulation of Starburst Dendrimers

The text of this section is part of an article coauthored with William A. Goddard III and Donald A. Tomalia. It is to be submitted to *Angewandte Chemie.*

Abstract

The structure and properties of two classes of starburst polymers were studied using molecular dynamics simulations.

One class, typified by the starburst dendrimers based on an ammonia core and β-alanine monomers, possesses open and extended forms for the early generations (one through three) but adopt more spheroidal structures containing internal hollows connected by channels that run the length of the assembly for generations 5 and above. The studies on these higher generation polymers indicate that: (a) ~50% of the surface area is *internal*, and (b) ~50% of the spheroidal volume is solvent-filled. Such systems provide the opportunity to selectively design unique polymers with internal binding sites. The sequestering of dopamine in a catechol-modified starburst dendrimer was modeled to illustrate the potential utility of these dendritic materials as therapeutic delivery systems.

A second class, typified by penta-erythritol polyether dendrimers, is found to have spherical forms at very early generations. For this type, the later generations lack any internal surface area or volume. Calculations of the amount of surface area available to terminal generation hydroxyl groups indicate the dense-packed limit for these polyether dendrimers to be the third generation, consistent with experiment.

1. Introduction

A novel class of polymers called "starburst dendrimers" has recently been discovered by Tomalia *et al.*[1-7] These unique polymeric materials (a) start with an initiator unit, termed the "core," which possesses multiple sites for condensation, and (b) use monomer subunits that terminate in functional groups, each of which allows for multiple branching sites. Using various synthetic strategies, it is possible to ensure that these polymers grow in a very systematic manner, producing materials with a well-defined number of monomer subunits and a quantized number of terminal groups, as shown schematically in Figure 1. Thus, the initial condensation of monomers to fully saturate the core unit produces a generation 1 dendrimer. The subsequent condensation of monomers to saturate the terminal functional groups of the generation 1 moiety yields the generation 2 dendrimer. This process is repeated to create higher generation polymeric dendrimers.

The strategies employed to synthesize this type of polymer involve the selective use of protecting groups[5] or multistep condensations[1,4] so that each generation can be completed before starting a new one. Thus nearly ideal, synthetically perfect materials can be made for characterization and can be used to proceed to subsequent generations. Unfortunately, even though each molecule has the same topological structure, there are an enormous number of possible conformations (these systems are fractal in nature) so that, despite their uniformity, there is no long-range order for x-ray crystal structure analysis. Thus, to date, there are no detailed experimental structural characterizations to guide the design of these new materials. This work (using molecular simulations) presents the first such structural characterization of starburst dendrimers. Based on this work, we are able to discuss some properties and possible functional applications.

Figure 1. Schematic two-dimensional representation of a starburst dendrimer growth cascade.

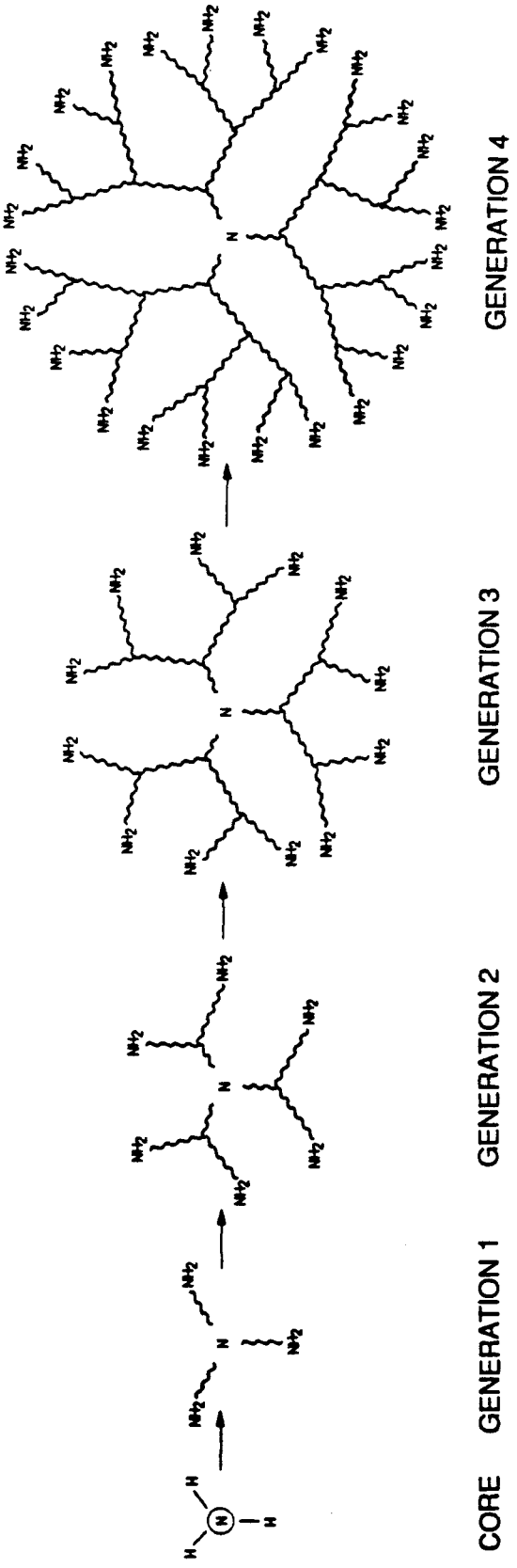


Figure 1.

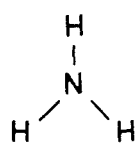
1.1 Synthetic Strategies

A variety of core units and monomer fragments have been used to synthesize a wide range of different starburst dendrimers containing various interior environments and surface functional groups. Generally speaking, these materials are produced via propagation techniques that include "time-sequenced, step-growth" components involving either: (a) differentiated, polyfunctional reagents used in molar excess quantities, or (b) elaborate protecting/deprotecting group chemistry. To assure starburst branching ideality, the condensations must be essentially quantitative, and the reaction conditions must be such that intramolecular looping or intermolecular bridging is minimized.

The molecular simulation studies presented in this work have focused on two types of starburst dendrimers: (a) the polyamidoamines and (b) the pentaerythritol polyethers. The synthetic methods to produce these materials and the experimental characterizations are summarized below.

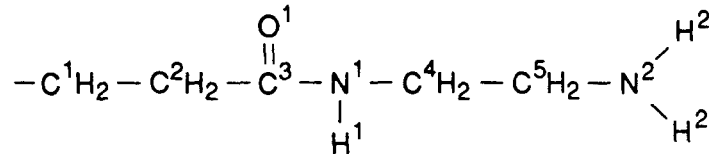
1.1.1 Polyamidoamine Starburst Dendrimers

The polyamidoamine dendrimers (PAMAM), also referred to as the β-alanine systems, begin with an ammonia core unit (**1**) and contain the amidoamine monomer building block (**2**) shown below.



CORE

1



MONOMER

2

The ammonia initiator core contains three condensation sites, thus producing a tridirectional dendrimer. Each monomer building block terminates with an

amino group containing two sites for branching and continued growth. Hence, the number of monomers added doubles from one generation to the next. After addition, these polymers have terminal, surface amino groups. Table 1 contains a listing of the number of monomers, per generation, for the PAMAM series.

The synthetic scheme employed in the production of these dendrimers falls into the excess reagent category. It is outlined in Figure 2. The first step in the synthesis involves exhaustive Michael addition of methylacrylate to a suitable amine group, either the core unit or the terminal amino groups from the preceding generation. The resulting esters are then exhaustively amidated with excess ethylenediamine to produce the next full generation polyamidoamine. Interruption of this synthetic cascade after the methylacrylate condensation yields ester-terminated oligomers, referred to as the half-generation materials.

Early work with the polyamidoamine starbursts indicated that it would be quite difficult to synthesize these highly branched polymers, without bridging, above generation 5.[6] Recent studies have shown, however, that generations one through ten can be produced highly monodispersed with high branching idealities ($\geq$ 95 mole %) as determined by C^{13} nuclear magnetic resonance spectroscopy.[8] Size exclusion chromatography (SEC), low-angle laser light scattering, and intrinsic viscosity measurements have been used to determine molecular weights, polydispersities, and hydrodynamic diameters of the polymers.[7] The results for generations 1 through 11 are listed in Table 2.

Table 1. Pertinent information for the β-alanine dendrimers investigated with molecular simulation.

Generation	Monomers		Condensation Sites	Total Atoms[a]
	New	Total		
1	3	3	6	34
2	6	9	12	94
3	12	21	24	214
4	24	45	48	454
5	48	93	96	934
6	96	189	192	1894
7	192	381	384	3814
8	384	765	768	7654
9	768	1533	1536	15334
10	1536	3069	3072	30694
11	3072	6141	6144	61414

[a] This is the number of atoms actually included in the simulation; i.e., it ignores hydrogen atoms bonded to carbon since these are treated implicitly.

Figure 2. Synthetic scheme for the polyaminoamide starburst dendrimers.

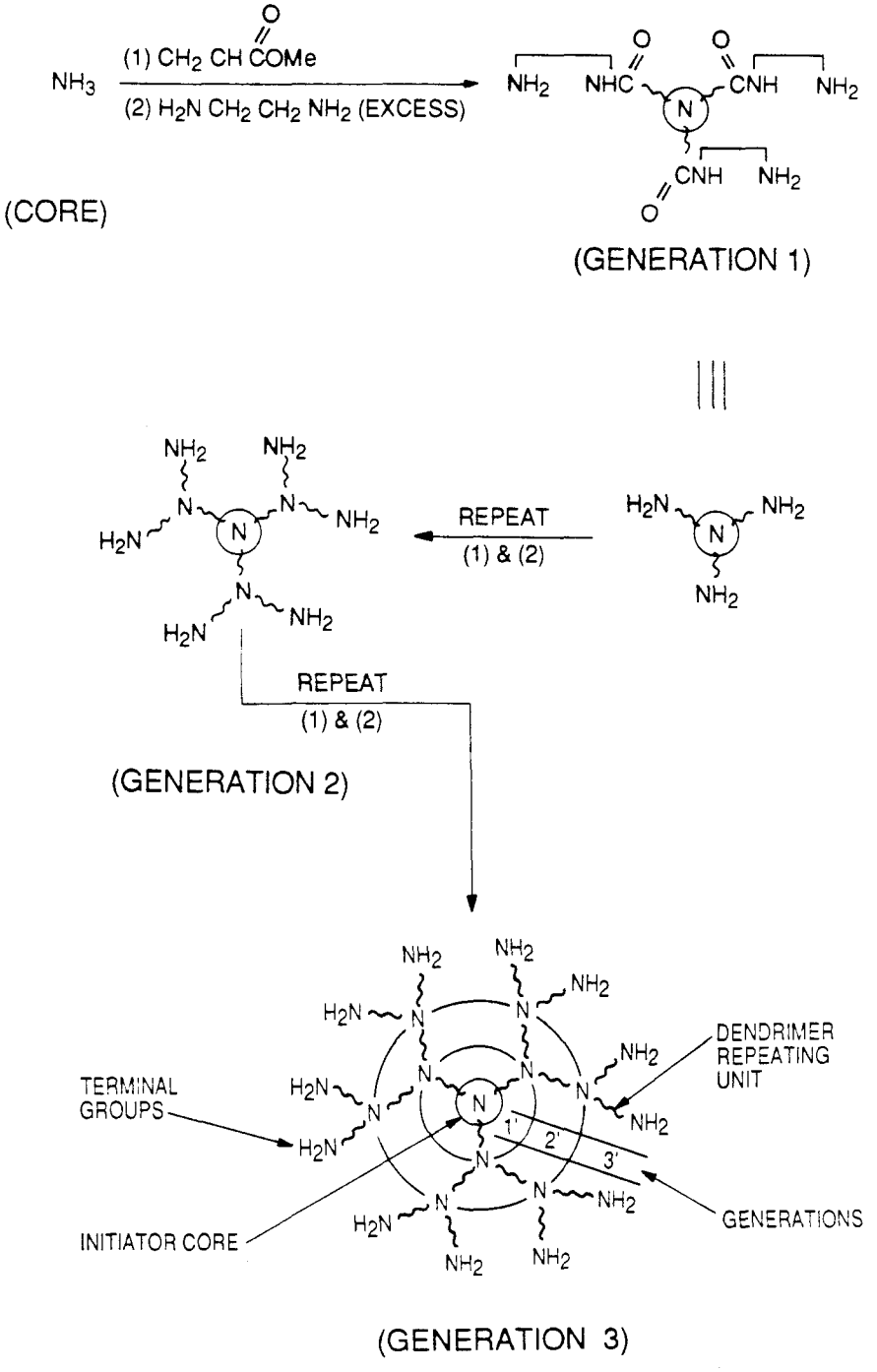

(1) CH2 CH COMe
NH3
(2) H2N CH2 CH2 NH2 (EXCESS)
(CORE)
(GENERATION 1)
REPEAT
(1) & (2)
(GENERATION 2)
TERMINAL GROUPS
DENDRIMER REPEATING UNIT
INITIATOR CORE
GENERATIONS
(GENERATION 3)

Figure 2.

Table 2. Experimentally determined dimensions for the dendrimers.

Generation	Molecular Weight		Diameter (Å)	
	Exp.[a]	Theory	Viscosity	S E C
PAMAM-1	-	360	11	11
PAMAM-2	-	1,044	17	16
PAMAM-3	-	2,070	22	24
PAMAM-4	5,200	5,154	31	31
PAMAM-5	10,700	10,633	40	40
PAMAM-6	21,600	21,591	53	56
PAMAM-7	44,000	43,503	67	66
PAMAM-8	88,000	87,341	76	75
PAMAM-9	174,000	175,006	90	90
PE-1				9.8
PE-2				18.6
PE-3				25.9
PE-4[b]				

[a] Molecular weight for generations 1 through 3 too low to be determined.

[b] Not experimentally available.

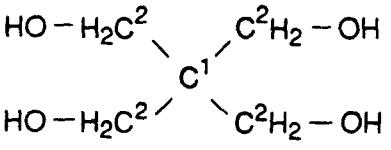
Table 3. Pertinent information for the polyether dendrimers investigated with molecular simulation.

Generation	Monomers		Condensation Sites	Total Atoms[a]
	New	Total		
1	4	5	12	53
2	12	17	36	173
3	36	53	108	533
4	108	161	324	1613

[a] This is the number of atoms actually included in the simulation; i.e., it ignores hydrogen atoms bonded to carbon since they are treated implicitly.

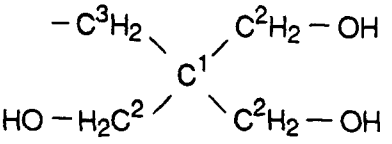
1.1.2 Polyether Starburst Dendrimers

The polyether dendrimers are composed of penta-erythritol fragments, both core (**3**) and monomer units (**4**), as depicted below.



CORE

3



MONOMER

4

The penta-erythritol core contains four hydroxyl condensation sites leading to a tetra-directional dendrimer. The monomer units are attached via hydrolytically stable ether linkages and terminate with three hydroxyl groups. Thus, the molecular weights of these polymers, as well as the number of surface hydroxyl groups, triple from generation to generation, as listed in Table 3.

The scheme employed in the synthesis of these macromolecules involves protect/deprotect strategies, as outlined in Figure 3. The branch junctures are preformed as part of a bicyclic orthoester reagent. Dendrimer construction involves using an alkoxide functionalized bicyclic orthoester to displace bromide groups from either : (a) penta-erythritol tetrabromide to produce the first generation, or (b) the appropriately brominated dendritic materials to yield higher generation materials.

Experimentally, these polyether oligomers have been synthesized and characterized up through generation 3. A complete fourth generation has been unattainable under present synthetic conditions. The available analytical data are presented in Table 2.

Figure 3. Strategy used for the synthesis of the penta-erythritol polyether starburst dendrimers.

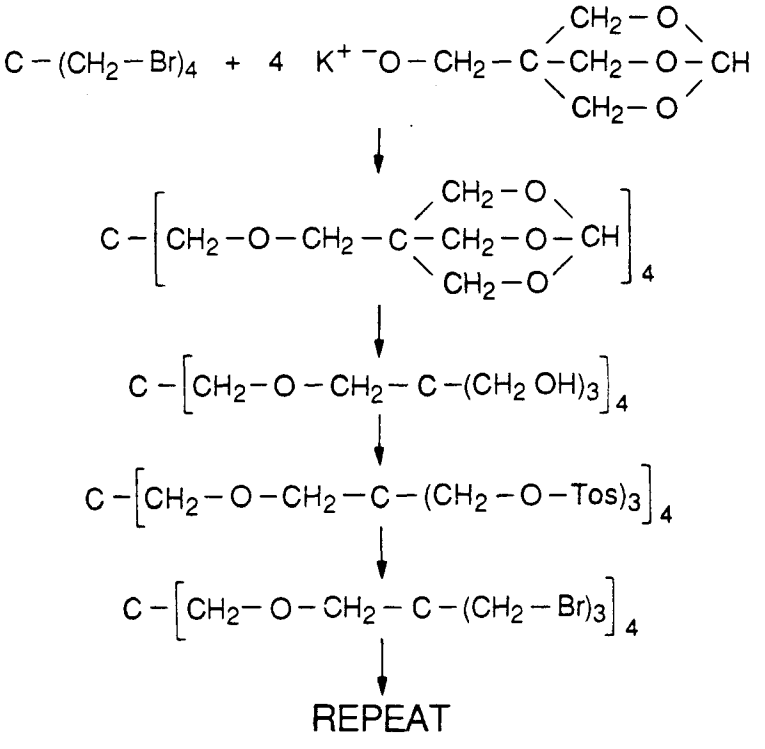

C – (CH2 – Br)4 + 4 K+ –O – CH2 – C – CH2 – O – CH
CH2 – O
CH2 – O
C – [CH2 – O – CH2 – C – CH2 – O – CH]4
CH2 – O
CH2 – O
C – [CH2 – O – CH2 – C – (CH2 OH)3]4
C – [CH2 – O – CH2 – C – (CH2 – O – Tos)3]4
C – [CH2 – O – CH2 – C – (CH2 – Br)3]4
REPEAT

Figure 3.

2. Calculational Details

2.1 Molecular Simulation Details

The dendrimers used as starting structures for the molecular dynamics simulations were "synthesized" using the polymer and dendrimer build facilities of POLYGRAF.[9] The molecular mechanics simulations were carried out with POLYGRAF using the AMBER force field developed for simulating proteins and nucleic acids.[10] The valence part of this force field describes the energy and forces in terms of a superposition of 2-body bond stretches, 3-body angle bends, and 4-body torsion and inversion terms. In addition, nonbond interactions are described with fixed charge electrostatic terms, with Lennard-Jones 12-6 potentials to account for van der Waals interactions, and a Lennard-Jones 12-10 potential to modulate the electrostatic terms involved in hydrogen bonding. A detailed description of the force field potential functions is presented in Appendix I.

For the electrostatic component of these simulations, partial charges for the monomer fragments were determined based upon values in AMBER[10] for analogous amino acid subunits. The partial charges used for each of the core and monomer subunits and the force field atom types are listed in Tables 4 and 5.

Initial molecular dynamics simulations were run for the β-alanine dendrimers generations 1 through 5 with explicit solvent present (within 10 Å of the polymer). For larger dendrimers we used a mean-field approximation to solvation. The common approximation with dielectric constant ϵ = R led to structures that were much more compact than when water molecules were included explicitly in the simulation. To obtain good agreement between mean-field and explicit solvent simulations required $\epsilon \approx 8R$. Hence, final simulations were run using a mean-field approximation for an aqueous environment that incorporated a distance-dependent dielectric constant with $\epsilon = 8R$.

Table 4. Force field atom types and point charge information for β-alanine dendrimers.

Atom[a]	Type[b]	IBA[c]	EBA[d]
C1	C_32	0.280	0.280
C2	C_32	0.039	0.039
C3	C_2	0.526	0.526
O1	O_2	-0.500	-0.500
N1	N_2H1	-0.520	-0.520
HN1	H_N	0.248	0.248
C4	C_32	0.200	0.054
C5	C_32	0.227	0.100
N2	N_3	-0.500	-0.867
HN2	H_N2	–	0.320

[a] See structure **2** for definitions.

[b] BIOGRAF Reference Manual.

[c] IBA = Inner β-Alanine.

[d] EBA = End β-Alanine.

Table 5. Force field atom types and point charge information for polyether dendrimers.

Atom[a]	Type[b]	CPE[c]	IPE[d]	EPE[e]
C1	C_3	0.400	0.256	0.256
C2	C_32	0.286	0.256	0.211
O	O_3	-0.386	-0.336	-0.550
H	H_O	–	–	0.31
C3	C_32	0.286	-0.016	-0.169

[a] See structures **3** and **4** for definitions.

[b] BIOGRAF Reference Manual.

[c] CPE = Core PolyEther.

[d] IPE = Inner PolyEther.

[e] EPE = End PolyEther.

The molecular dynamics calculations were run at a simulation temperature of 300°K and time step of 1 fs with structural and energetic information recorded at 50 fs sampling intervals. For structure and data analysis, the first picosecond was used strictly for equilibration of the system, and the remaining simulation time was used for data collection and analysis.

2.2 Determination of Properties

Various structural properties of these macromolecular assemblies were calculated from the molecular simulation trajectories. The average end-to-end distance, average radius of gyration, and average principal moments of inertia were determined from the final 4 ps of the simulations. Various molecular surfaces were generated for a structure for each dendrimer generation.

2.2.1 End-to-end Distances

The simplest measure of the "size" of the macromolecule is the maximal end-to-end distance, which is defined as the average of the *maximal* terminal heteroatom-to-terminal heteroatom distance determined at the end of each of the 800 sampling intervals. This may be considered as the diameter, as discussed further in Section 3. For the β-alanine series, this diameter is the maximum terminal nitrogen-to-terminal nitrogen distance, while in the polyether class, the distances are maximal terminal oxygen-to-terminal oxygen positions.

2.2.2 Radius of Gyration

Another measure of the size of a macromolecule is given by its radius of gyration, which can be thought of as the average displacement of each atom from the center of mass (cm) of the system as defined in (1).

$$R_g^2 = \frac{1}{M_{Tot}} \sum_{\text{all atoms}} m_i[(x_i - x_{cm})^2 + (y_i - y_{cm})^2 + (z_i - z_{cm})^2] \qquad (1)$$

Here, M_{Tot} is the total mass and number of atoms. The R_g value was calculated at the end of each sampling interval of the molecular simulation, and the values from the final 4 ps were averaged. For a sphere of uniformly distributed atoms, the sphere diameter (D_g) is related to R_g by $D_g = 2\sqrt{5/3} \cdot R_g$

2.2.3 Moments of Inertia

The moments of inertia provide a simple measure of molecular shape. The moment of inertia tensor is defined in (2).

$$I_{xy} = \frac{1}{M_{Tot}} \sum_{\text{all atoms}} m_i(x_i - x_{cm})(y_i - y_{cm}) \qquad (2)$$

Diagonalizing $\{I_{ij}\}$ leads to the principal axes and principal moments I_x, I_y, and I_z. If $I_x = I_y = I_z$ the molecule is spherical. If $I_z \approx 0$ and $I_x = I_y$, the molecule is linear. If $I_x = I_y$ and $I_z \neq 0$, the molecule is a symmetric top that is prolate if $I_z < I_{x,y}$ or oblate if $I_z > I_{x,y}$. The values for the total moment of inertia and the components along the principal axes were calculated from the simulated polymer structures at each sampling interval and averaged over the final 4 ps.

2.2.4 Surface Calculations

Richards[11] introduced two useful rigorous quantitative definitions for the qualitative concept of a molecular surface: (1) the vdw *molecular surface* and (2) the solvent *accessible surface*. Consider that each atom is surrounded by a sphere corresponding to its van der Waals radius. The vdw molecular surface is the lotus of exposed sphere segments. The solvent accessible surface is produced by rolling a sphere with a radius of the solvent (1.4 Å for water) around the vdw surface. The solvent accessible surface is the locus of sphere midpoints, and hence an internal exposed surface will be found only for cavities sufficiently large to hold the solvent.

In order to obtain quantitative measure for the internal cavities as a function of polymer growth, we calculated the molecular surfaces for the dendrimers for a sequence of probe radii from 0 to 11 Å.

2.2.4.1 Molecular Surfaces

The vdw molecular surface is made up of two portions: (a) the contact surface, which is that part of the van der Waals surface readily accessible to the probe sphere and (b) the re-entrant surface, defined as the inward facing portion of the probe sphere when it is in contact with more than one atom. The molecular surface maps out the *contact points* between the probe sphere and the atomic van der Waals spheres as illustrated in Figure 4. Thus, surfaces generated with larger and larger probe spheres become more smooth and approach a limiting surface as the probe radius approaches infinity.

An algorithm for generating vdw molecular surfaces was used in the display and analysis of the starburst dendrimers. The method uses the approach developed by Michael Connolly[12] and is available in POLYGRAF. All molecular surfaces were calculated with a density of 8.0 dots per $Å^2$. For generations 5 and above (in the β-alanine set), the number of probes necessary for the surfacing exceeded the dimensions of the program. To circumvent this problem the surfaces were calculated in segments, determined by generation and excluding those atoms of different generations, patched together, and then corrected for the slight discrepancy in the area associated with each patch.

Figure 4. Schematic representation of molecular surface definitions. A section through part of the van der Waals envelope of a dendrimer monomer is shown with the atomic centers numbered. The vdw molecular surface is illustrated by the heavy line surrounding the atomic spheres. The accessible surface is mapped out by the center of probe spheres 1 (solid line) and 2 (dashed line).

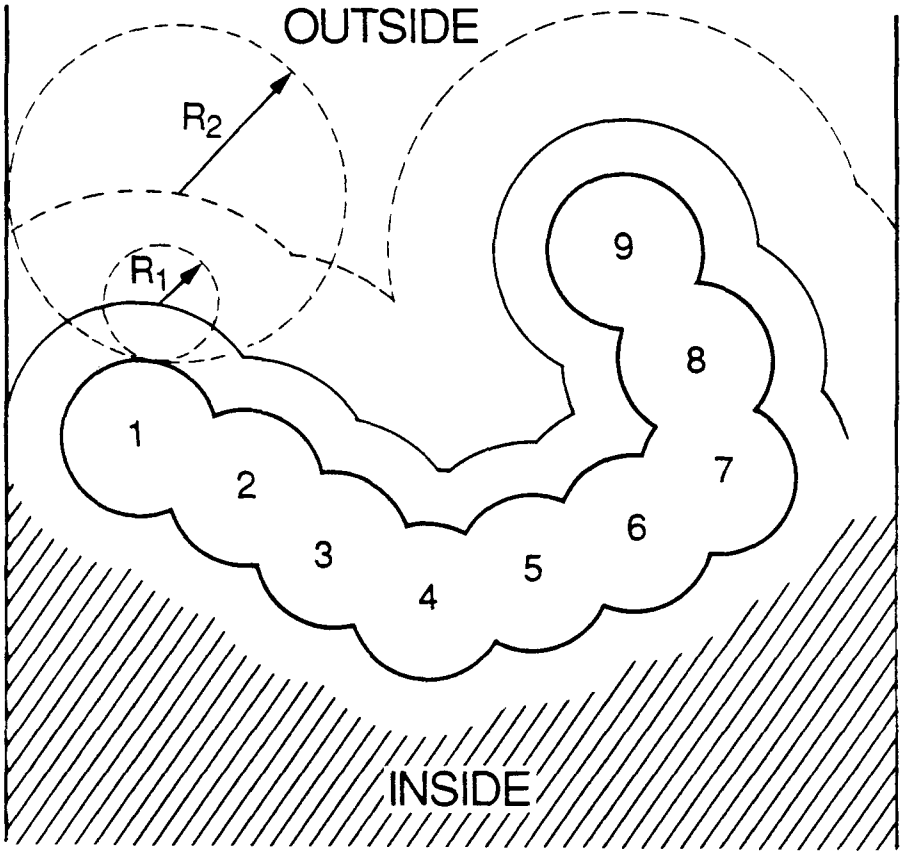


Figure 4.

2.2.4.2 Accessible Surfaces

The *accessible surface* of a molecule differs from its molecular surface counterpart in that the surface is defined in terms of the *center* of the probe sphere as it is rolled along the van der Waals' surface of the molecule. Thus, the accessible surface produced continues to increase in size as the radius of the probe sphere increases. This surface type is also illustrated in Figure 4.

The GEPOL87 Program,[13] available from the Quantum Chemistry Program Exchange but not yet interfaced with a graphics facility, was the method used to calculate the solvent accessible surfaces. This algorithm is a modification of the Connolly methodology, which incorporates a means of estimating the volume contained within the surface, as described in 2.2.5. The NDIV parameter was set to 3 for all calculations.[14] The dimensions of this program are such that the higher generation polyamidoamine polymers could be treated as one entire entity.

2.2.5 Molecular Volumes

GEPOL87,[13] in the process of creating the solvent accessible molecular surfaces, uses a polyhedral procedure to determine appropriately sized polyhedrons surrounding each atom. The volume enclosed by such polyhedral surfaces is then calculated and tallied to produce an estimate of the total volume contained within the surface. Thus, we have also generated the volume information for those systems surfaced with GEPOL87.

3. Results and Discussion

The molecular dynamics simulations show that the character of the structures adopted by the β-alanine dendrimers changes as the number of generations increases. The early generation dendrimers (generations 1 through 3) are very open and well extended, resembling starfish with arms shooting out in various directions. These are just large molecules and have a hemispherical disk structure, with a curvature arising from the nitrogen core and with no real inside to the polymer. Wire-frame structures of generations 1 through 4 are shown in Figure 5. This open, extended topology for the early generation systems is strikingly different from earlier proposals,[3] but these hemispheric structures provide a simple interpretation of observed nuclear magnetic resonance relaxation times[15] (preprint in Appendix II). The molecular simulations indicate a dramatic change in the overall structural properties for the β-alanine systems as they grow past generation 4. What is starfishlike for the early generation PAMAM's becomes a tangled mass by generation 5. At higher generations the structures are more spherelike, as illustrated in Figure 6, and the later generations do tend to be on the outside of the polymer.

A better definition of the structure is given in Figures 7 and 8 where the vdw molecular surface is displayed, thus defining the outside of the molecule. The early generations are somewhat amorphous in shape, but the later ones are clearly shperical. These views from the outside give little clue as to the internal character of these systems. The internal structure is shown more clearly in Figures 9 and 10, where graphical slabbing is used to cut out a section from the middle of the polymer. The early generation dendrimers have little internal character but already by generation 4 there are substantial nooks and crannies, which by generation 6 have evolved into a labyrinth of channels and cavities.

Figure 5. Wire-frame presentations of the starfishlike early generation β-alanine starburst dendrimers (one through four) . The core nitrogen is depicted in red, the terminal generation is shown in light blue, and the inner generations are yellow.

Figure 6. Wire-frame presentations of the later generation β-alanine starburst dendrimers (five through seven) . The core nitrogen is depicted in red, the terminal generation is shown in light blue, and the inner generations are yellow. Generation 4 is also displayed to facilitate size comparisons with the earlier generation polymers, as presented in Figure 5.

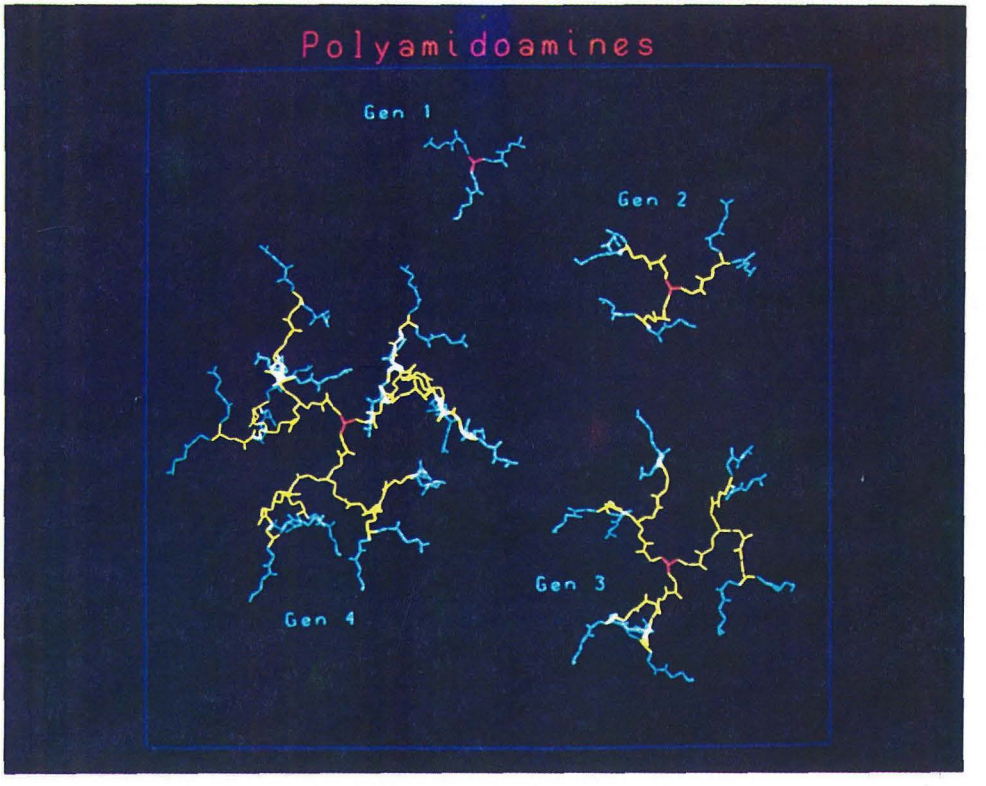


Figure 5.

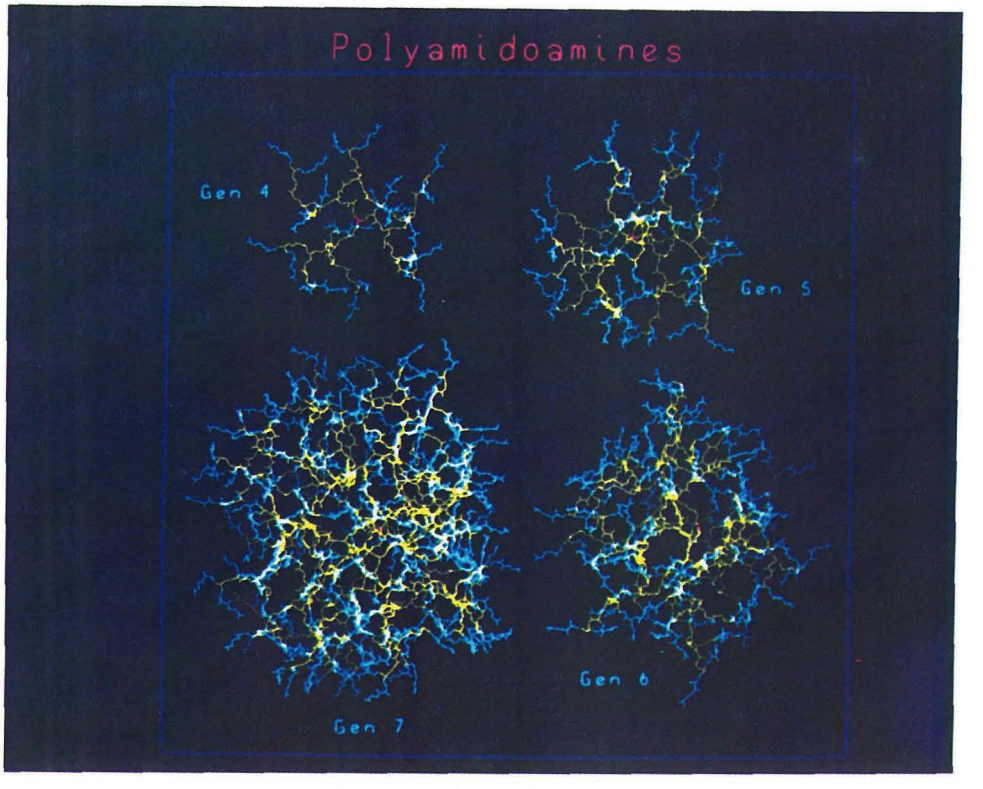


Figure 6.

Figure 7. Vdw molecular surface displays for the early generation β-alanine starburst dendrimers (one through four) illustrating the varied shapes adopted by these open, starfishlike molecules.

Figure 8. Vdw molecular surface displays for the later generation β-alanine starburst dendrimers (five through seven) illustrating the spheroidal nature of these macromolecules. Generation 4 is also included to illustrate its position as transition generation and to facilitate shape comparisons with the earlier generation polymers, as presented in Figure7.

Figure 7.

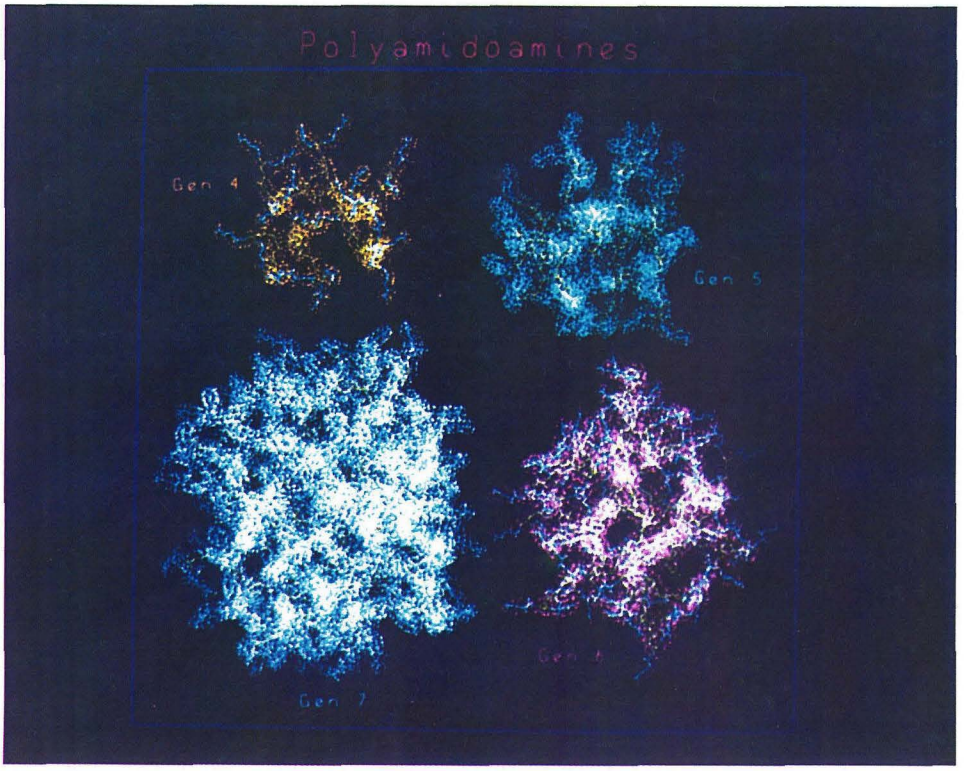


Figure 8.

Figure 9. Slabbed views of the vdw molecular surface displays for the early generation β-alanine starburst dendrimers (one through four) illustrating the lack of an interior region to these polymers.

Figure 10. Slabbed views of the vdw molecular surface displays for the later generation β-alanine starburst dendrimers (five through seven) illustrating the extensive internal channels and cavities associated with these materials.

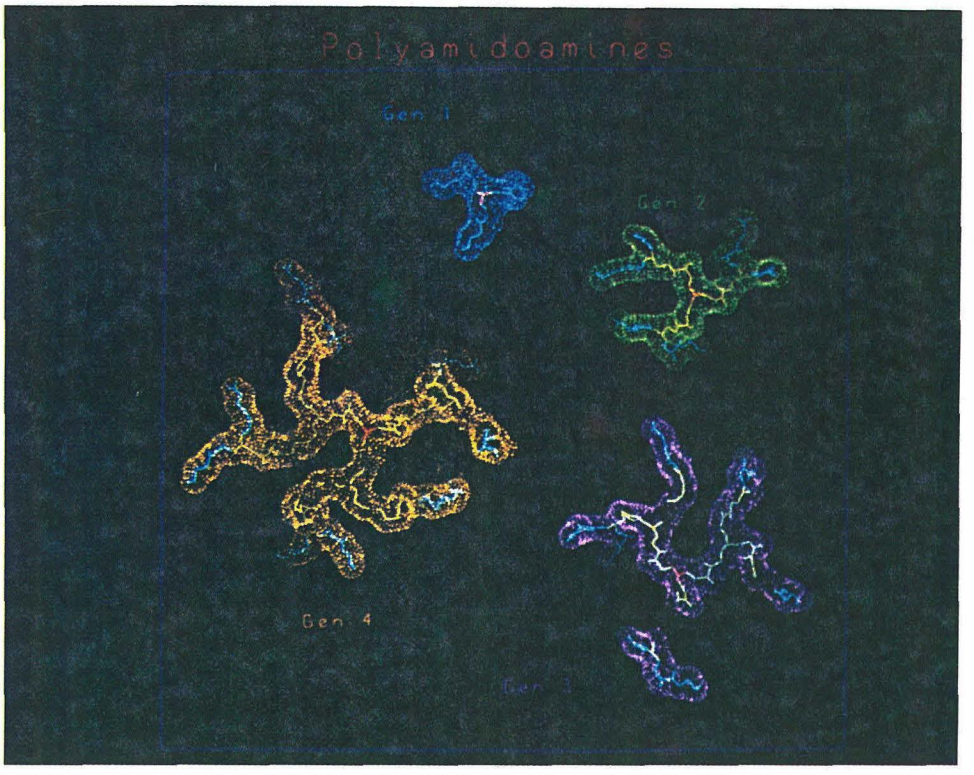


Figure 9.

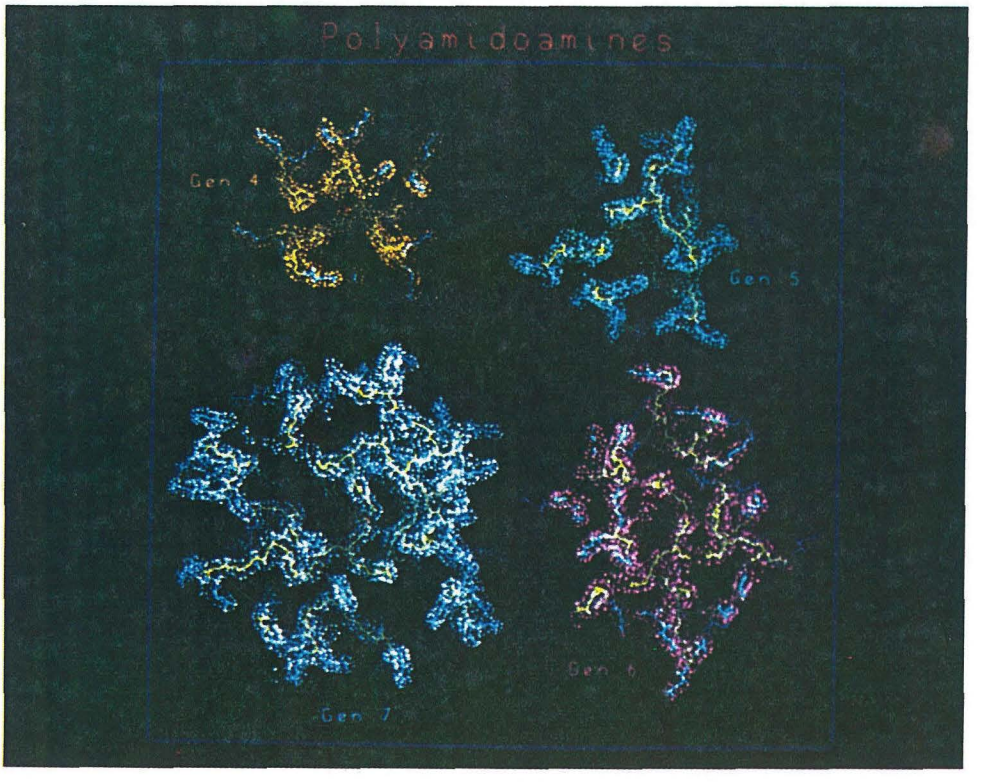


Figure 10.

These later systems clearly have well-defined internal structures.

In contrast to PAMAM, the penta-erythritol polyether dendrimers show dramatically more compact, spherical structures by the third generation. Generation 1 might be loosely considered open and extended, but at generation 2, and certainly above, these polymers are very densely packed spheroidal moieties lacking in any appreciable interior void spaces. This is shown in Figure 11 (wireframe), Figure 12 (full vdw surface), and Figure 13 (interior slabs). The addition of the vdw molecular surface, as depicted in Figure 13, clearly shows the densely packed characteristics of the interior of these materials.

In order to quantify these structural characteristics, a series of computational analyses were performed, as summarized below.

3.1 End-to-end Distances

The average diameter for both sets of polymers increases linearly with generation. Table 6 contains the calculated maximal end-to-end distances for generations 1 through 7 of the β-alanine dendrimers. We find that these diameters range from $\sim$13Å (for generation 1) through $\sim$80Å (for generation 7) with an increase of $\sim$10Å per generation.

Previously, Tomalia *et al.* used CPK models as a means of estimating limits to the sizes of the β-alanine systems.[3] These diameters were obtained by: (1) fully extending a CPK model to measure an upper limit to the molecular diameter and (2) compressing the CPK unit to arrive at a lower diameter limit. These CPK-derived dimensions do indeed bracket the simulated maximal end-to-end distances, as shown in Figure 14. A comparison of the calculated diameters with those determined via viscosity and size-exclusion chromatography (SEC) studies (given in Table 2) indicates that the maximal end-to-end distances are $\sim$10Å

Figure 11. Wire-frame presentations of the penta-erythritol, polyether starburst dendrimers (generations one through four). The core penta-erythritol unit is depicted in red, the terminal generation is shown in yellow, and the inner generations are drawn in light blue.

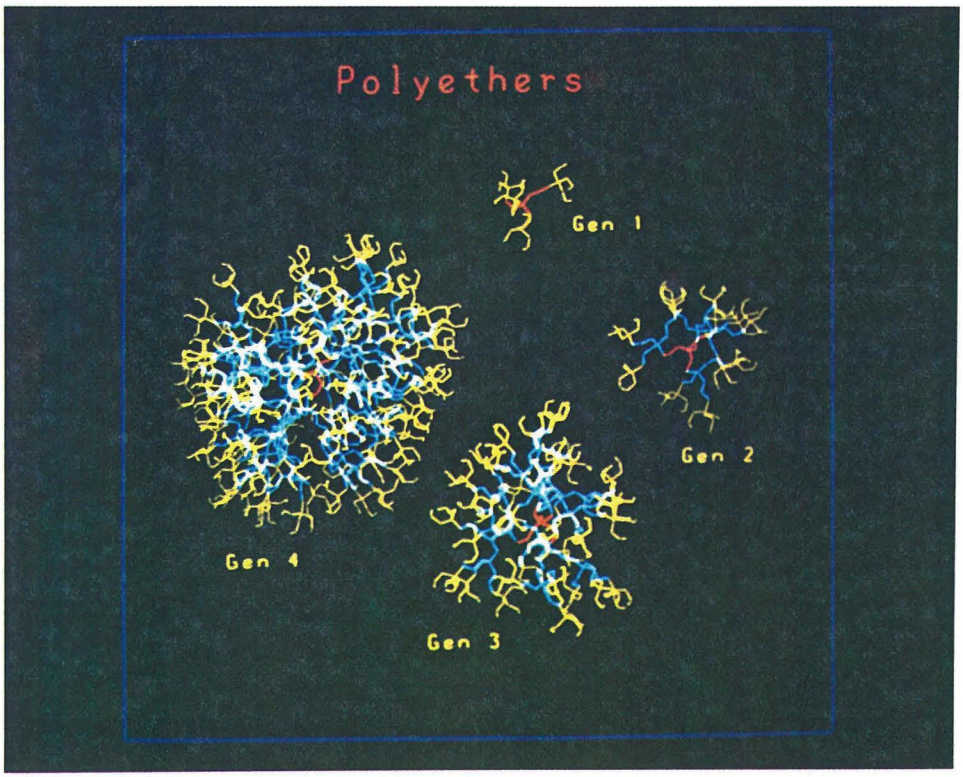


Figure 11.

Figure 12. Vdw molecular surface displays for the penta-erythritol, polyether starburst dendrimers (generations one through four) illustrating the dense, spheroidal nature of these macromolecules.

Figure 13. Slabbed views of the vdw molecular surface displays for the penta-erythritol polyether starburst dendrimers (generations one through four) illustrating the congestion and lack of extensive, internal, void spaces associated with these materials.

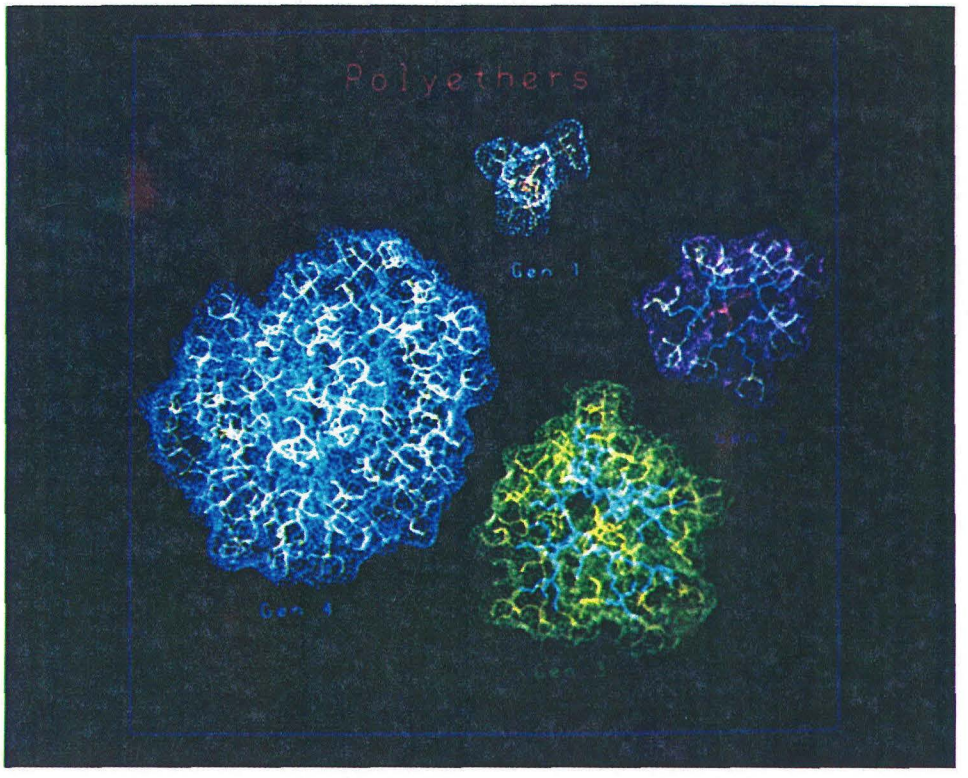


Figure 12.

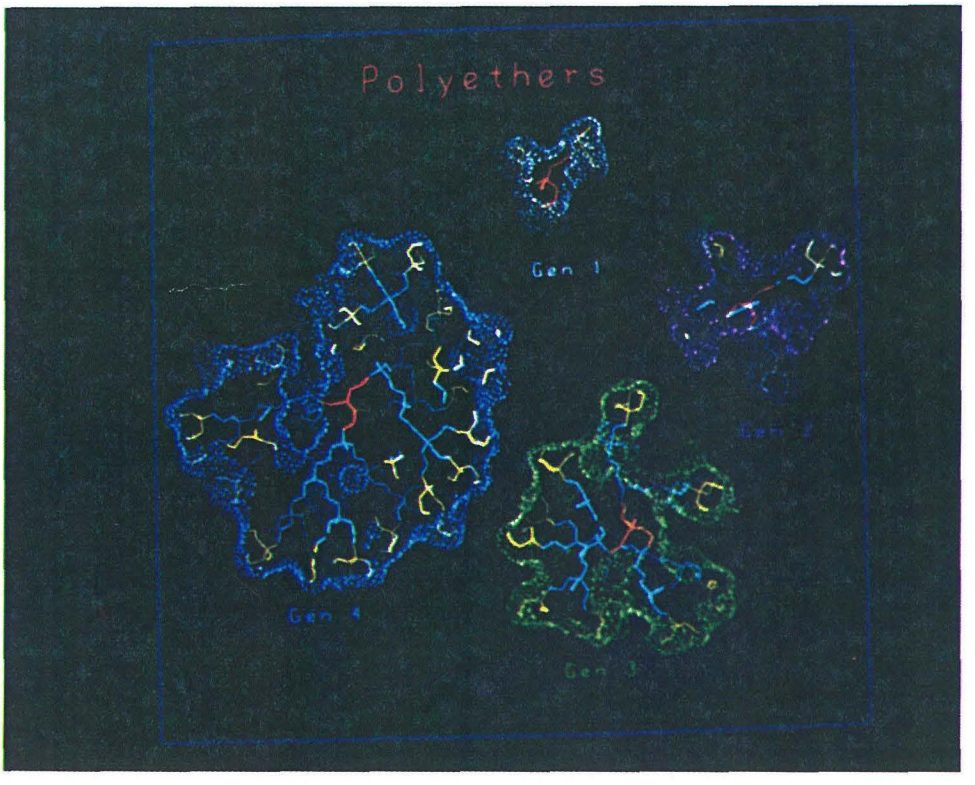


Figure 13.

Table 6. Simulated maximal end-to-end distance and radius of gyration for the dendrimers.

Generation	End-to-end Distance (Å)	Radius of Gyration (Å)	
		R_g	$D^a_{R_g}$
PAMAM-1	12.63	5.77	14.9
PAMAM-2	24.33	9.81	25.3
PAMAM-3	34.55	13.47	34.8
PAMAM-4	47.03	15.82	40.9
PAMAM-5	56.53	18.90	48.8
PAMAM-6	68.47	21.52	55.6
PAMAM-7	79.85	26.91	69.5
PE-1	12.69	5.13	13.3
PE-2	21.40	7.76	20.0
PE-3	29.93	10.51	27.1
PE-4	40.06	14.43	37.3

[a] D_{R_g} defined in Section 2.2.2.

Figure 14. Diameters for the PAMAM starburst dendrimers. The CPK-based values (dotted lines) bracket both the experimental (solid line) and simulated (dashed lines) dimensions.

Figure 15. Diameters for the polyether starburst dendrimers. The CPK-based values (dotted lines) fall short of the experimental numbers (solid line), while the simulated dimensions (dashed line) are a bit high.

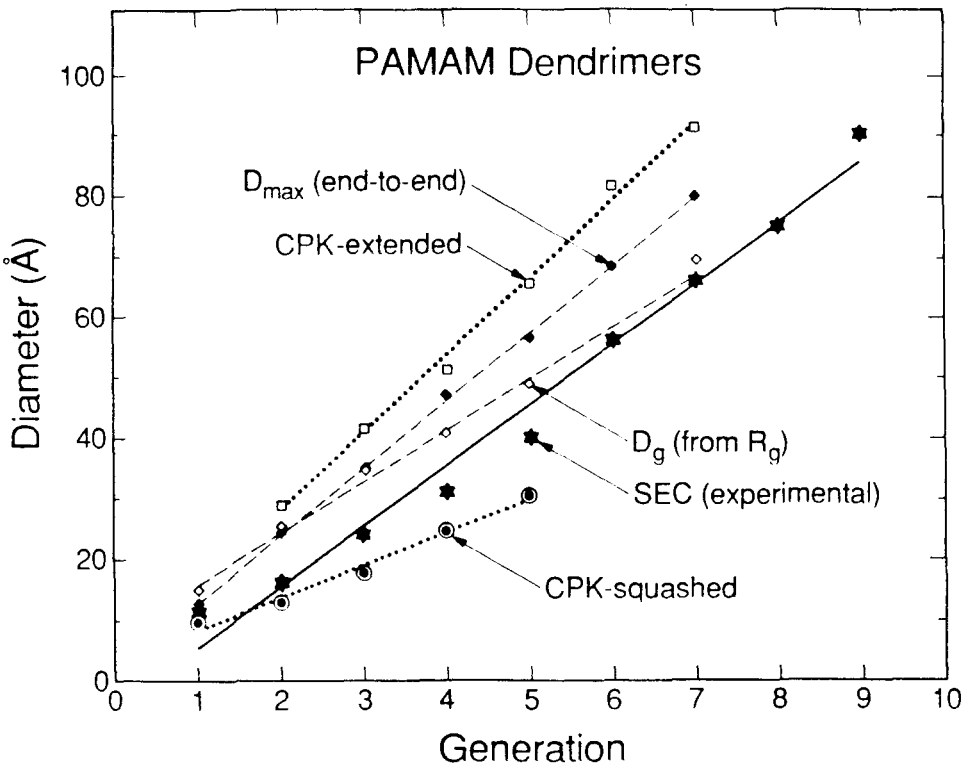


Figure 14.

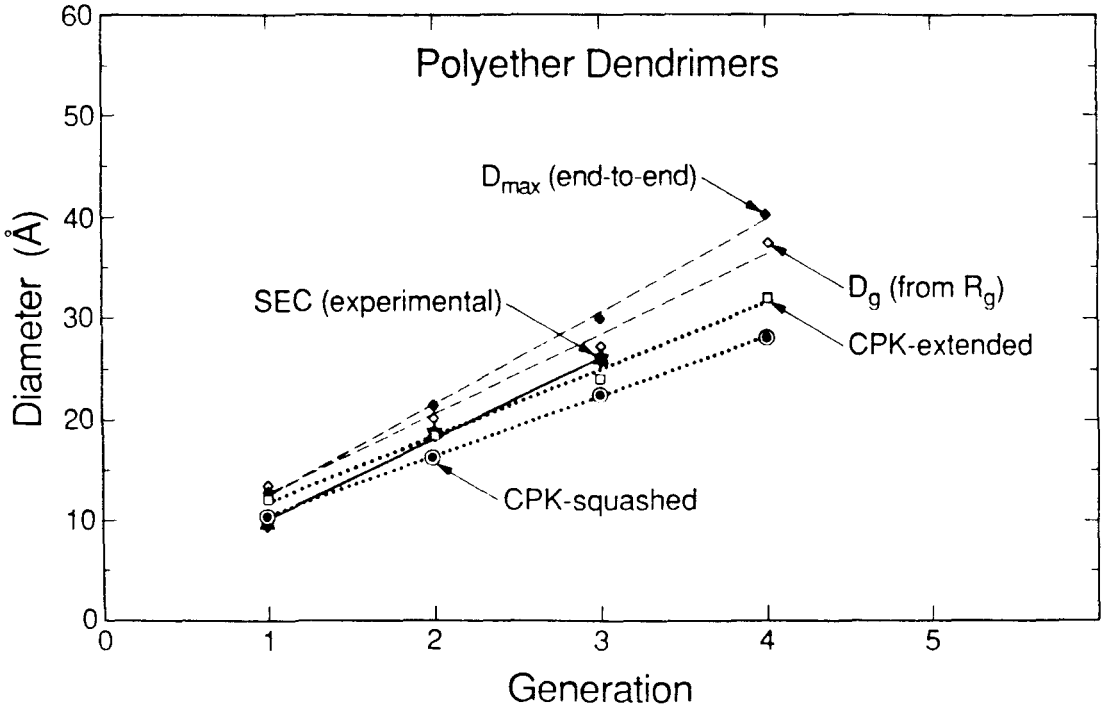


Figure 15.

larger than the experimental diameters. This difference is most likely due to the simulated values, taking into account *only* the maximal end-group displacement. Averaging many across-polymer distances may lead to a more representative diameter for these systems.

The polyether dendrimers also exhibit a linear increase in average maximal end-to-end distance as reported in Table 6. Because of the differences in length and flexibility of the penta-erythritol units relative to the β-alanine fragment, these polymers span a ~13Å to ~40Å range (generations 1 to 4) with an average increase at each generation of ~9Å. The calculated values are in very good agreement with estimates for the diameters from SEC measurements. A plot of these simulated dimensions, plus the CPK-based values reported by Tomalia and co-workers, is presented in Figure 15. In this case, the simulated values exceed those derived from CPK models for the higher generation polyethers. This difference is most probably due to the simulated structures' alleviating some of the congestion by allowing some slight distortions in the bond and angular components of the system. The CPK units, on the other hand, are locked into idealized sp^3 hybridized angles and bond lengths.

3.2 Radius of Gyration

The radius of gyration of a macromolecule is also indicative of its size. The calculated radius of gyration, R_g, and its corresponding spherical diameter for the investigated dendrimers are presented in Table 6. Within each of the series studied, the radius of gyration increases with generation. For the higher generation β-alanine dendrimers as well as the later generation polyethers, the diameters derived from the radius of gyration are in excellent agreement with the experimental diameters. The predicted diameters for the low generation dendrimers

are a bit high when compared with the experimental data. This difference is due to the nonspherical nature of the early generation dendrimers (as discussed below). The experimental measurements would reflect the minimal "diameter" of the polymer ellipsoid and hence would be smaller than a uniformly spherical diameter.

An examination of the radius of gyration of these macromolecules as a function of the total number of dendrimer monomers is shown in Figure 16. These plots indicate that the early generation β-alanines exhibit a linear increase in R_g as a function of the monomer number, which is indicative of the open, extended form adopted by these molecules. The higher generations, however, show a more gradual increase in R_g with the number of monomers, deviating from the rapid rise in R_g displayed by the early generation samples. This behavior is consistent with the spheroidal form adopted by the larger polyamine dendrimers. For the polyether series, the change in R_g as a function of the total number of monomer units is gradually increasing in an exponential fashion, indicative of a spherical shape.

By comparing equivalent generations from the two sets, we see that the penta-erythritol-based dendrimers have a radius of gyration that is smaller than the corresponding β-alanine macromolecule. For instance, the generation 3 polyether value has R_g $\sim$10.5, while the PAA generation 3 polymer has R_g $\sim$13.5. Thus, the radius of gyration measure indicates that the polyether macromolecules are more compact than their polyaminoamide counterparts.

Figure 16. The calculated radius of gyration versus the total number of monomers for the PAMAM and polyether starburst dendrimers.

Figure 17. The aspect ratio (I_z/I_x) versus generation for the PAMAM and polyether starburst dendrimers.

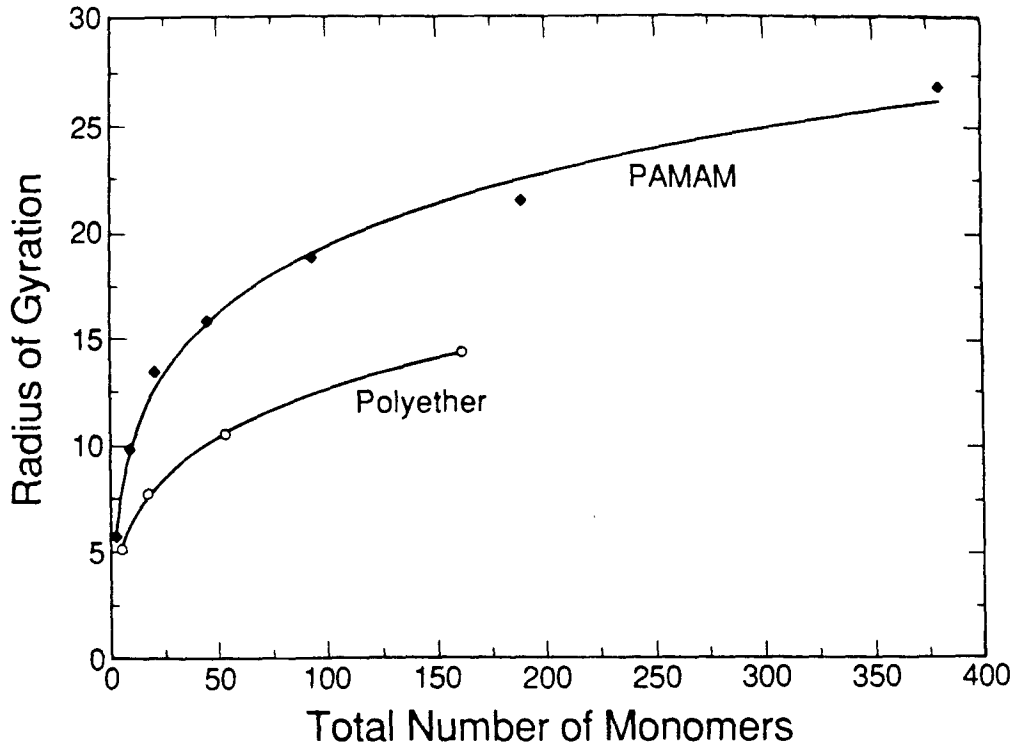


Figure 16.

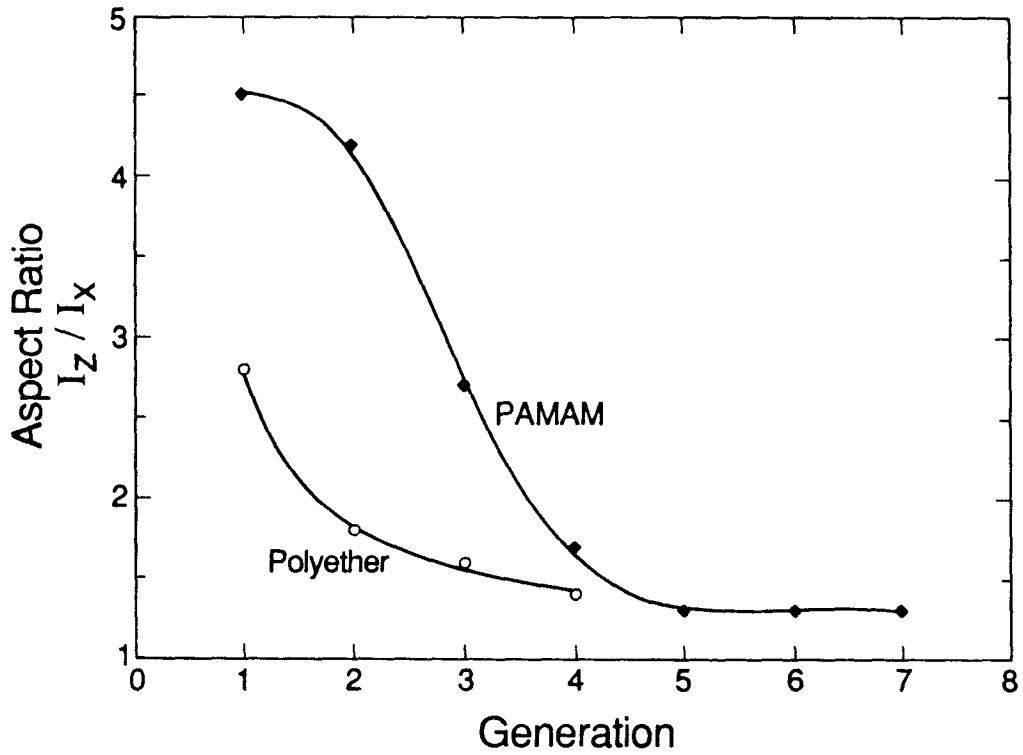


Figure 17.

3.3 Moments of Inertia

To quantify the changes in shape of the macromolecules with generation, the total moments of inertia, as well as their components along the principal axes, are reported in Table 7.

For the β-alanine series, the asymmetry of the structures clearly decreases as the number of monomers in the polymer increases. The early generations are quite asymmetric (Iz/Ix = 4.4 to 2.7), with all three principal moments unequal. The asymmetry drops to 1.7 for generation 4 and continues to decrease to a plateau at 1.3 for generations 5 and above. A plot indicating this change in asymmetry is presented in Figure 17. These components of the moment of inertia indicate that above generation 4 the PAMAM polymers have the form of an oblate spheroid.

The polyether dendrimers, on the other hand, have principal moments that are more evenly distributed at quite an early stage of dendrimer growth (see Figure 17). Even at generation 2 the aspect ratio for this series has dropped to 1.8 and by generation 4 it is 1.4. Hence, at generations 2 and above, these penta-erythritol polymers have a more regularized, spheroidal shape.

3.4 Surface Calculations

To quantify the changes in structure associated with dendrimer growth, e.g., the open and accessible surface area of the polymer and the amount of interior void space present, solvent accessible surfaces for the two series of starburst dendrimers have been calculated at different probe radii. By varying the size of the probe, it is possible to determine the dependence of surface area and volume on probe size and then estimate the amount of interior surface area and internal volume available for contact with guest and/or solvent molecules.

Table 7. Moment of inertia data for the dendrimers.

Generation	I^a_{Total}	%I_X	%I_Y	%I_Z	I_Z / I_X
PAMAM-1	1.056	11.26	38.58	50.16	4.5
PAMAM-2	9.914	13.85	28.61	57.54	4.2
PAMAM-3	43.24	18.77	29.92	51.31	2.7
PAMAM-4	129.3	24.59	34.85	41.56	1.7
PAMAM-5	379.2	29.06	34.54	36.40	1.3
PAMAM-6	991.6	29.52	33.54	36.94	1.3
PAMAM-7	3131.	29.88	31.16	38.96	1.3
PE-1	1.459	17.09	35.82	47.09	2.8
PE-2	11.43	23.38	35.03	41.59	1.8
PE-3	65.38	26.71	31.12	42.17	1.6
PE-4	381.5	28.53	32.51	38.96	1.4

[a] Units of 10^4amu Å^2.

3.4.1 Total Accessible Surfaces

The area of the accessible surfaces, calculated at various probe radii for the β-alanine and polyether dendrimers, are reported in Tables 8 and 9.

General

In order to guide the analysis of the solvent accessible surfaces, we consider an ideal spherical case. For such a system, the surface area, S_{AS}, is defined as

$$S_{AS} = 4\pi(R + p)^2, \tag{3}$$

where R is the radius of the ideal sphere and p is the radius of the probe used to calculate the accessible surface. Rearranging this equation to arrive at the dependence of S_{AS} on p produces:

$$\sqrt{S_{AS}} = 2\sqrt{\pi}R + 2\sqrt{\pi}p. \tag{4}$$

For this idealized case, the slope of the plot is 3.545, and the p=0 intercept is at $\sqrt{S_{AS}} = 3.545R$. Thus, the intercept of a linear plot of p vs $\sqrt{S_{AS}}$ provides a means of determining the radius of the spherical system.

Polyether

Graphs of $\sqrt{S_{AS}}$ vs. p for the polyether starbursts are shown in Figure 18. We find that for $p \geq 4.2$ Å, the dependence of $\sqrt{S_{AS}}$ on p is indeed linear. However, the surface area is larger at smaller values of p. This deviation is attributed to internal curvature and surface irregularities. Linear regression analysis[16] was performed on various subsets of the data to determine which set of points was most appropriate for a linear fit. The results indicate that using those values with $p \geq 4.2$ Å yield linear fits with correlation coefficients of 1.000 and standard errors in intercept and slope less than 0.22 and 0.03, respectively. (Adding data

Table 8. Accessible surface area ($Å^2$) for the polyether dendrimers calculated with GEPOL87.

	Probe Radius			
Generation	0.0	1.4	2.8	4.2
1	549.17	765.54	1051.26	1398.53
2	1781.85	1934.31	2235.70	2667.28
3	5148.33	3900.64	3873.04	4370.07
4	14368.73	6845.72	6610.91	7152.19
	Probe Radius			
Generation	5.6	7.0	8.4	9.8
1	1796.87	2256.94	2754.01	3281.44
2	3196.05	3774.21	4387.94	5102.32
3	4992.10	5684.79	6446.97	7211.77
4	7889.57	8725.35	9584.28	10511.3

Table 9. Accessible surface area (Å^2) for the β-alanine dendrimers calculated with GEPOL87.

	Probe Radius			
Generation	0.0	1.4	2.8	4.2
1	358.50	656.32	980.99	1334.14
2	1013.68	1726.50	2259.79	2748.47
3	2373.64	3910.87	4977.63	5624.37
4	5011.30	7726.73	9117.60	9766.34
5	10236.64	14699.83	16117.63	15283.05
6	21252.73	28028.39	25933.17	21868.78
7	42952.36	54285.37	44796.43	34105.96
	Probe Radius			
Generation	5.6	7.0	8.4	9.8
1	1731.22	2183.89	2690.10	3239.48
2	3268.57	3843.61	4503.23	5203.55
3	6194.16	6865.55	7641.02	8441.37
4	10349.21	11106.71	11970.14	12889.87
5	15150.64	15577.47	16415.86	17367.34
6	21042.03	21646.39	22432.90	23540.87
7	30214.66	29851.21	30811.59	31779.59
	Probe Radius			
Generation	11.2	12.6	14.0	
1	3833.08	4463.65	5148.77	
2	5959.41	6752.39	7619.38	
3	9349.68	10307.94	11332.52	
4	13870.10	14969.33	16134.53	
5	18491.75	19746.39	20971.44	
6	24895.35	26230.37	27647.16	
7	33183.39	34738.20	36346.64	

Figure 18. The $\sqrt{\text{accessible surface area}}$ versus probe radius for the polyether starburst dendrimers. Solid lines represent linear regression fits ($p \geq 4.2$), while dashed lines connect the calculated values.

Figure 19. The $\sqrt{\text{accessible surface area}}$ versus probe radius for the PAMAM starburst dendrimers. Solid lines represent linear regression fits ($p \geq 7.0$), while dashed lines connect the calculated values.

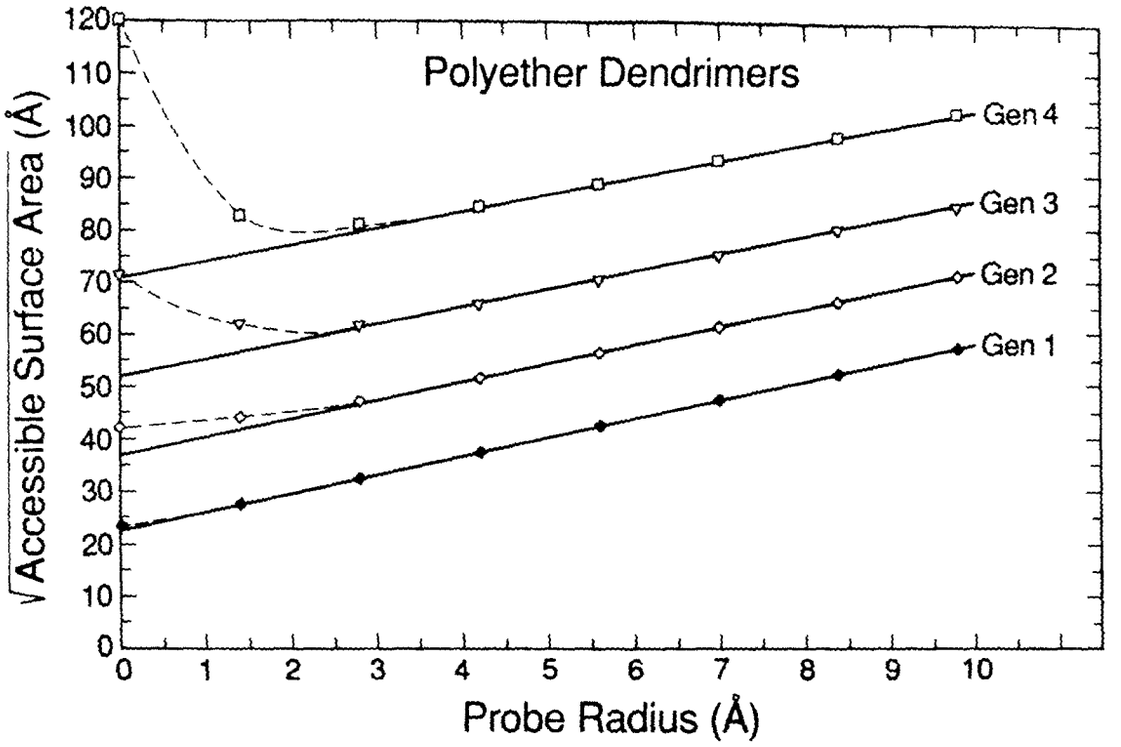


Figure 18.

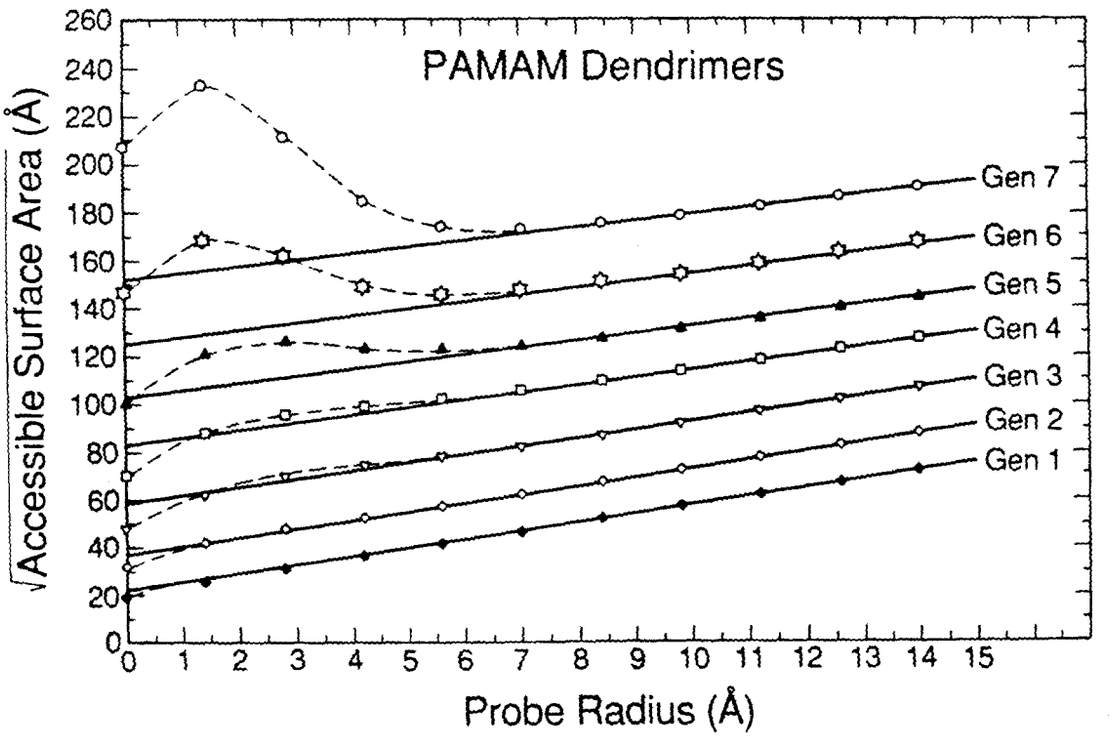


Figure 19.

points with p less than 4.2 gave fits with much higher error values.) The results from fitting these data to the equation

$$\sqrt{S_{AS}} = B_0 + B_1 \cdot p \tag{5}$$

are reported in Table 10. Once again, the polyether structures behave as spheroidal units. The slopes (B_1) range from 3.6 to 3.2, all close to the value $B_1 = 3.545$ for a sphere. The sphere radii calculated from the intercepts compare quite well with the diameter calculated from the radius of gyration and the experimental size-exclusion chromatographic measurements (see Tables 2 and 6).

The deviation from linearity observed at small p leads to valuable additional information. A probe with large p cannot interact with any interior regions of the polymer, and hence, the surface produced reflects purely the exterior portion of the macromolecule. For smaller probe spheres, however, the probe can access certain interior regions of the molecule and can produce an accessible surface that contains segments of both the exterior and the interior regions of the macromolecule. Indeed, this deviation from linearity at small p enables one to determine what portion of the calculated accessible surface is due to interior hollows. This value is arrived at using: (a) the regression equation to predict the surface area expected of a spheroidal unit at $p = 1.4$ Å; (b) subtracting this amount from the calculated accessible surface area (again with $p = 1.4$ Å); and (c) taking the ratio of this difference to the expected external accessible surface area. For the polyether polymers, no more than ~20% of the accessible surface area is associated with interior regions of the molecule.

PAMAM

The $\sqrt{S_{AS}}$ versus p analysis of the PAMAM series shows that the higher generation dendrimers, i.e., 4 and above, have profiles similar to the polyether

Table 10. Regression fits of probe radius versus (accessible surface area)$^{\frac{1}{2}}$ for the polyether dendrimers.

Generation	B_0^a	B_1^a	R_{Ideal}^b	$R_{B_1}^c$	Reg^d	GEPOL^e	Excess^f
1	22.487	3.561	6.34	6.32	754.7	765.5	0.01
	(0.165)	(0.023)					
2	36.830	3.519	10.39	10.47	1744.	1934.	0.11
	(0.186)	(0.026)					
3	51.851	3.375	14.63	15.36	3201.	3901.	0.22
	(0.161)	(0.022)					
4	70.946	3.214	20.01	22.07	5709.	6846.	0.20
	(0.214)	(0.029)					

[a] The value given in parenthesis represents the standard error.

[b] Calculated using ideal slope of 3.545.

[c] Calculated using fit slope, B_1.

[d] Calculated from the regression fit with probe = 1.4 Å.

[e] Calculated using GEPOL87[13] with probe = 1.4 Å.

[f] Fraction of surface that is internal.

macromolecules (see Figure 19 and Table 11). With larger probe spheres the accessible surfaces mapped out do depend linearly upon p. For the β-alanines, the best regression fits were achieved with $p \geq 8.4$ Å. The correlation coefficients were all ≥ 0.997. For this series, however, the slopes of the fitted lines are $\sim$3.0 instead of the 3.545 value expected of a perfect sphere. The radii of the molecules determined from the intercepts again agree quite well with the D_g obtained from the predicted radius of gyration (see Table 6).

The amount of estimated internal surface area increases dramatically above generation 4. For the generation 4 dendrimer, which we have already concluded (on the basis of the aspect ratio analysis) to be at the transition from an open and extended structure to a more spheroidal moiety, only $\sim$2% excess accessible surface area is found (see Table 11). By generation 7, however, we find that the internal surface is 120% of the external surface area. This provides a quantitative measure of the internal hollows so obvious in Figure 10. For generations 5 and 6, the internal surface is 30% and 70%, respectively, of the amount of external surface area. Thus, from generations 5 thru 7 the interior surface area is increasing and there is no indication of the filling-in expected when surface congestion dominates.

3.4.2 Normalized Accessible Surface Area

The total accessible surface area data (presented in Tables 8 and 9) may also be used to calculate the change in surface area per added monomer, i.e., the normalized accessible surface area. This analysis of the data is given in Tables 12 and 13. Figures 20 and 21 illustrate this normalized change in surface area plotted as a function of increasing probe size.

Table 11. Regression fits of probe radius versus (accessible surface area)$^{\frac{1}{2}}$ for the β-alanine dendrimers.

Generation	B_0^a	B_1^a	R_{Ideal}^b	$R_{B_1}^c$	Reg^d	GEPOL^e	Excess^f
1	22.13	3.55	6.24	6.23	734.1	656.3	-0.11
	(0.118)	(0.010)					
2	36.87	3.60	10.40	10.24	1756.	1727.	-0.02
	(0.096)	(0.008)					
3	58.61	3.41	16.53	17.19	4018.	3911.	-0.03
	(0.382)	(0.034)					
4	82.784	3.146	23.35	26.31	7602.	7727.00	0.02
	(0.578)	(0.051)					
5	102.542	3.009	28.93	34.08	11400	14700	0.29
	(0.739)	(0.065)					
6	124.636	2.965	35.16	42.04	16590	28030	0.69
	(0.569)	(0.050)					
7	151.918	2.739	42.86	55.47	24260	54290	1.24
	(1.453)	(0.128)					

[a] The value given in parenthesis represents the standard error.

[b] Calculated using ideal slope of 3.545.

[c] Calculated using fit slope, B_1.

[d] Calculated from the regression fit with probe = 1.4 Å.

[e] Calculated using GEPOL87[13] with probe = 1.4 Å.

[f] Fraction of surface that is internal.

Table 12. Accessible surface area ($Å^2$) per added monomer for the polyether dendrimers calculated with GEPOL87.

	Probe Radius			
Transition	0.0	1.4	2.8	4.2
Core → 1	99.84	119.89	142.28	167.91
1 → 2	102.72	97.40	98.70	105.73
2 → 3	93.51	54.62	45.48	47.30
3 → 4	85.37	27.27	25.35	25.76
	Probe Radius			
Transition	5.6	7.0	8.4	9.8
Core → 1	194.62	222.57	249.52	266.46
1 → 2	116.60	126.44	136.16	151.74
2 → 3	49.89	53.07	57.20	58.60
3 → 4	26.83	28.15	29.05	30.55

Table 13. Accessible surface area (Å^2) per added monomer for the β-alanine dendrimers calculated with GEPOL87.

	Probe Radius			
Transition	0.0	1.4	2.8	4.2
Core → 1	104.97	173.74	234.87	289.79
1 → 2	109.19	178.36	213.13	235.72
2 → 3	113.33	182.03	226.49	239.66
3 → 4	109.90	158.99	172.50	172.58
4 → 5	108.86	145.27	145.83	114.93
5 → 6	114.75	138.84	102.25	68.60
6 → 7	113.02	136.76	98.25	63.74
	Probe Radius			
Transition	5.6	7.0	8.4	9.8
Core → 1	342.68	395.72	456.05	508.39
1 → 2	256.23	276.62	302.19	327.35
2 → 3	243.80	251.83	261.48	269.82
3 → 4	173.13	176.72	180.38	185.35
4 → 5	100.03	93.14	92.62	93.28
5 → 6	61.37	63.22	62.68	64.31
6 → 7	47.77	42.73	43.64	42.91
	Probe Radius			
Transition	11.2	12.6	14.0	
Core → 1	562.11	611.82	660.16	
1 → 2	354.39	381.46	411.77	
2 → 3	282.52	296.30	309.43	
3 → 4	188.35	194.22	200.08	
4 → 5	96.28	99.52	100.77	
5 → 6	66.70	67.54	69.54	
6 → 7	43.17	44.31	45.31	

Figure 20. The normalized surface area versus accessible surface probe radius for the polyether dendrimers. Dashed lines connect the calculated values.

Figure 21. The normalized surface area versus accessible surface probe radius for the PAMAM dendrimers. Dashed lines connect the calculated values.

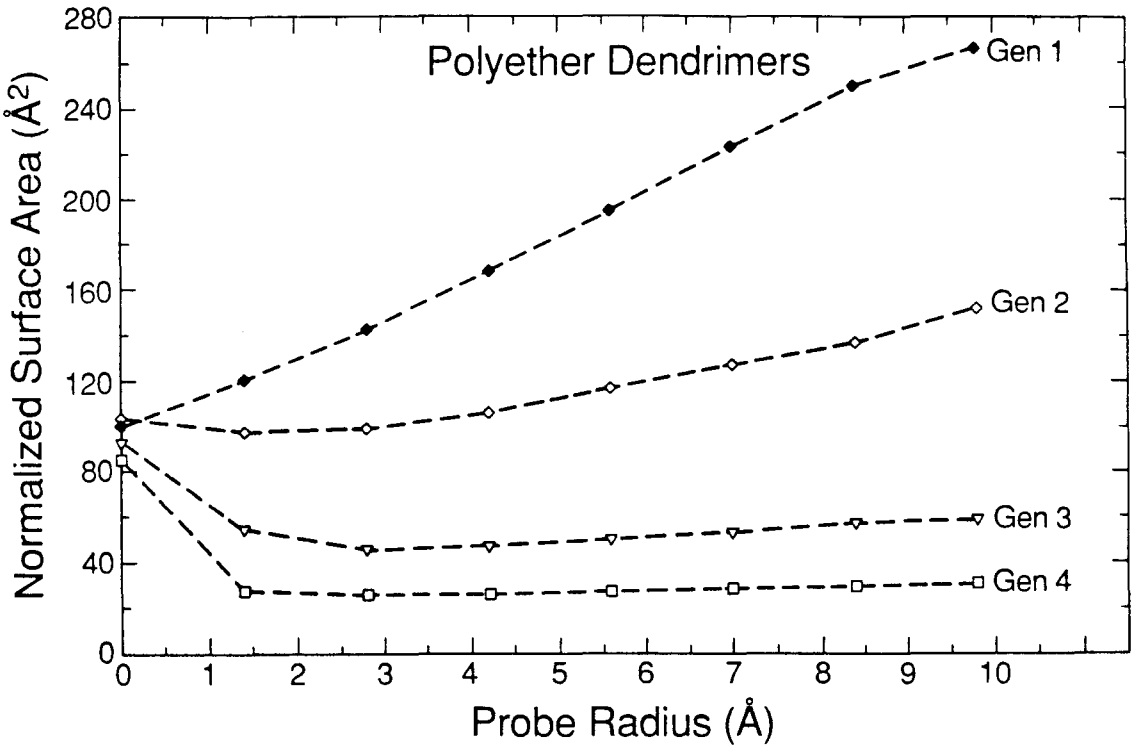


Figure 20.

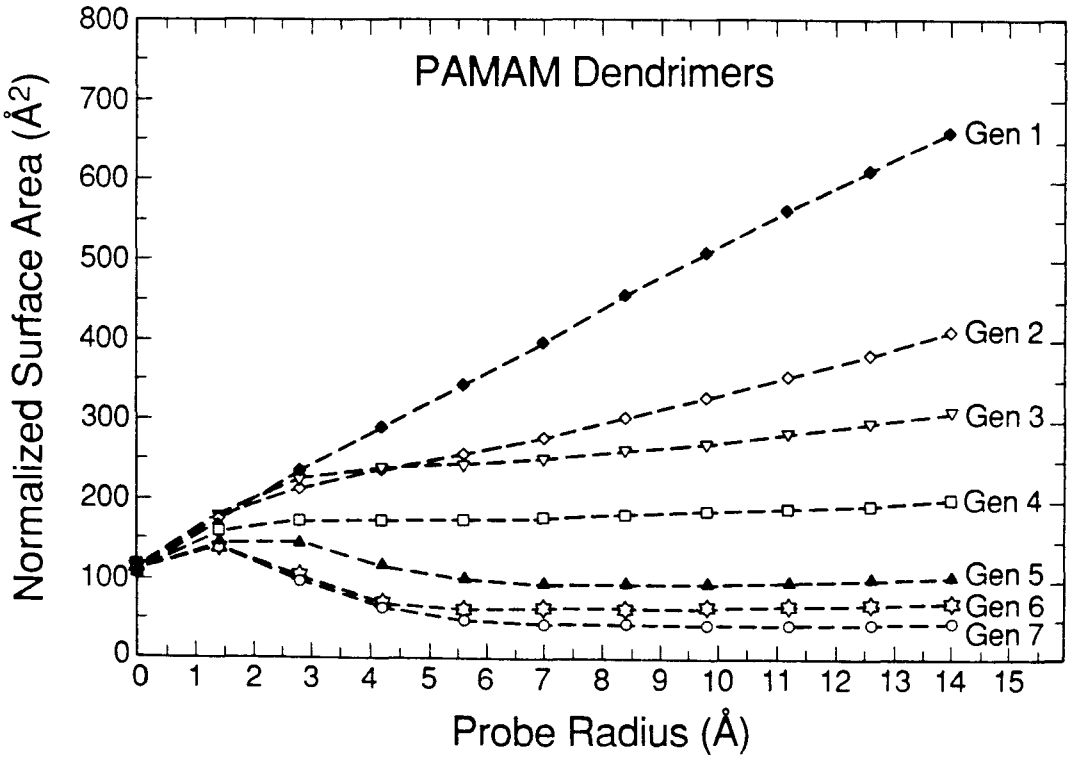


Figure 21.

Polyether

For the polyether dendrimers, the van der Waals area (probe = 0.0 Å) is approximately constant for all four cases. As the size of the probe grows, however, only generations 1 and 2 permit the accessible surface area to increase with the probe sphere. Both generations 3 and 4 exhibit enough congestion that the area per monomer cannot rise with the value of the probe.

ΔS_{AS} per monomer at $p = 1.4$ provides a means of calculating the area per hydroxyl head group on the surface of the polymer sphere. Each terminal hydroxyl group has access to $\sim 40 Å^2$ at generation 1, $\sim 30 Å^2$ at generation 2, $\sim 20 Å^2$ at generation 3, but $< 10 Å^2$ by the fourth generation. This area is significantly less than the area measured for dense-packed hydroxyl groups in a bilayer plane,[17] indicating the sterically forbidden nature of this fourth generation polyether.

PAMAM

For the β-alanine polymers, the early generations (one through four) are open enough to allow for an increase in the normalized surface area as the measuring probe is increased. At the fifth generation and above, however, the spheroidal nature of the polymers requires that monomers from the last added generation be near enough to each other that the change in surface area begins to decrease with increasing probe sphere. At the seventh generation, $\sim 150 Å^2$ are still available to each terminal amino group, well above the dense-packed limit.

3.5 Volume Calculations

The calculated volumes contained within the solvent accessible surfaces allow estimates of the volume contained within interior regions of the polymer matrices. For the polyether series, the above discussions suggest that there will be little

or no internal volume. On the other hand, such analyses should offer a means of quantifying the amount of open, solvent-filled space available to the higher generation β-alanine dendrimers.

3.5.1 Total Accessible Surface Volumes

We again return to the sphere as our model for deriving accessible surface relationships. For this system, the volume enclosed by the accessible surface V_{AS} would be:

$$V_{AS} = \frac{4}{3}\pi(R + p)^3, \tag{6}$$

where R is the radius of the ideal sphere and p is the radius of the probe used to calculate the surface. Rearranging Equation 6 to arrive at the dependence of V_{AS} on p yields:

$$\sqrt[3]{V_{AS}} = \sqrt[3]{\frac{4}{3}\pi} \cdot R + \sqrt[3]{\frac{4}{3}\pi} \cdot p. \tag{7}$$

For this ideal case, the slope of the plot is 1.612 while the intercept is 1.612 R. As with the analysis of the surface areas of the accessible surfaces, the intercept of a linear plot of p vs $\sqrt[3]{V_{AS}}$ provides another means of estimating the radius of the spherical system.

Plots of $\sqrt[3]{V_{AS}}$ as a function of probe radius for both the polyether and polyamidoamine dendrimers are shown in Figures 22 and 23. As in the surface area analysis, those volumes calculated with large values of p show a linear dependence on probe radius. A deviation from this linearity is observed for the volumes calculated with smaller probe spheres. Now the volume is less than ideal because of internal cavities.

Polyether

For the polyether macromolecules, the solvent accessible volume data are

Figure 22. The $\sqrt[3]{\text{accessible volume}}$ versus probe radius for the polyether starburst dendrimers. Solid lines represent linear regression fits ($p \geq 2.8$), while dashed lines connect the calculated values.

Figure 23. The $\sqrt[3]{\text{accessible volume}}$ versus probe radius for the PAMAM starburst dendrimers. Solid lines represent linear regression fits ($p \geq 5.6$), while dashed lines connect the calculated values.

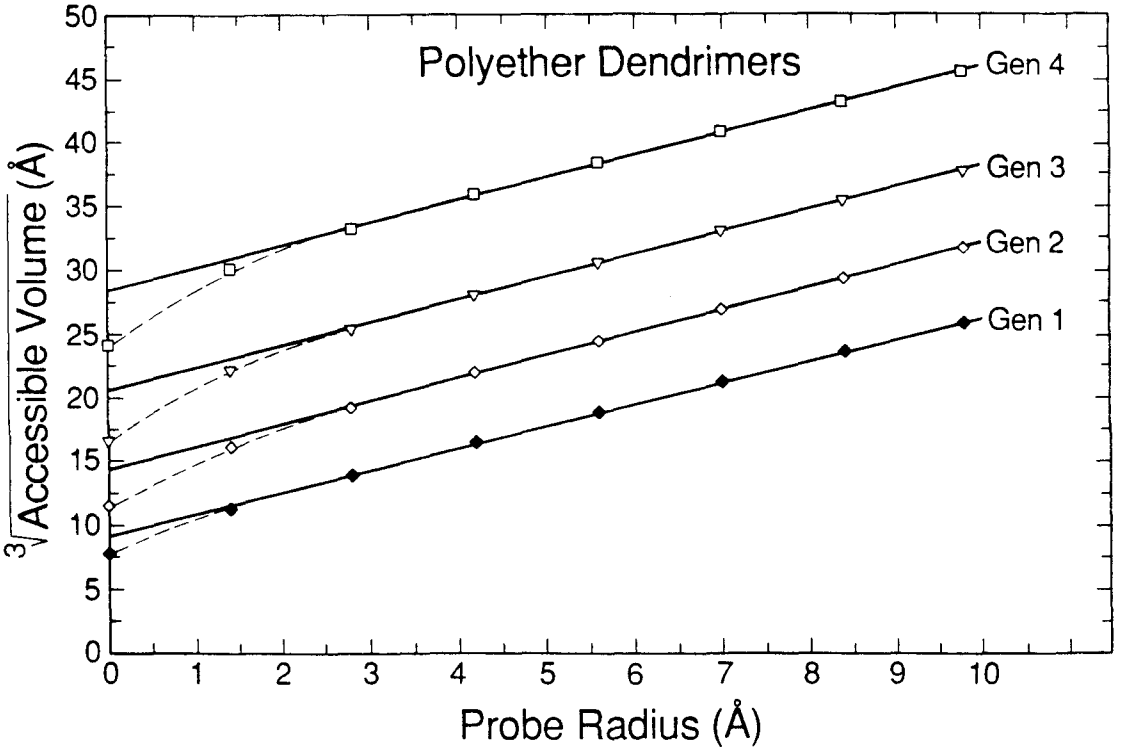


Figure 22.

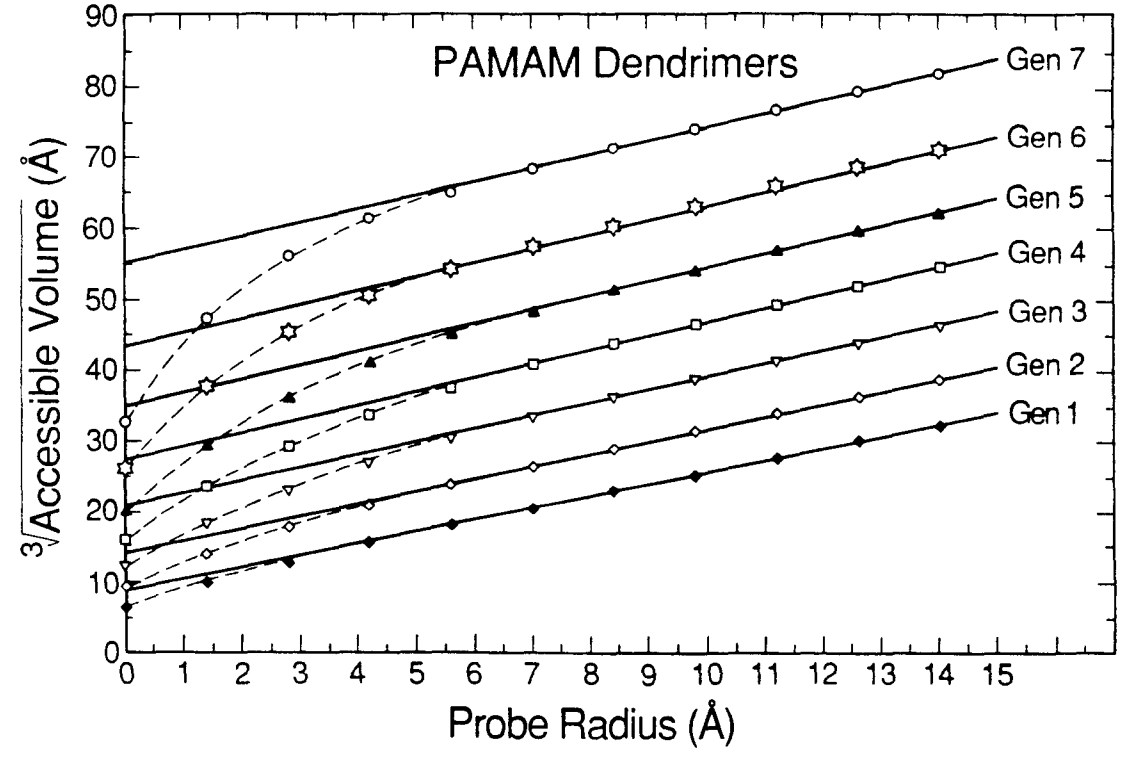


Figure 23.

given in Table 14. We find an excellent fit of the data to the equation

$$\sqrt[3]{V_{AS}} = B_0 + B_1 \cdot p \qquad (8)$$

(regression analysis on those data points with $p \geq 2.8$ Å produces fits with correlation coefficients of 1.000 and standard errors in intercept and slope less than 0.20 and 0.025, respectively). The slopes, B_1 ≈1.7, are slightly greater than the value of 1.612 derived for the perfect sphere, while the radii calculated based upon the intercepts are in very good agreement with the SEC values and only 1-2 Å shorter than both those values derived from the surface area analysis and the simulation end-to-end distance averages.

These regression equations can be used in conjunction with the volume calculated at p=1.4Å, to estimate the amount of internal volume found in each generation of these highly branched polymeric materials. Based on the ratio C/D, where C is the difference between the volume predicted from the regression equation and that calculated directly with GEPOL and D is the predicted value of the volume (all at p = 1.4 Å), we conclude that less than 15% of the volume available to the polyether dendrimers is composed of open, solvent-filled void spaces. These values are reported in Table 15.

PAMAM

For the higher generation polyamidoamine series (data presented in Tables 16 and 17), we find a good fit to Equation 8 for large p (using $p \geq 7.0$ Å leads to correlation coefficients of 1.00 and standard errors less than 0.30 and 0.03 in intercept and slope, respectively). The values for the regression coefficients are given in Table 28. These models indicate that the higher generation β-alanine macromolecules deviate from perfect spheroids much more so than do the polyether molecules. The regression slopes are ~1.97, as opposed to the

Table 14. Volume (Å^3) for the polyether dendrimers' accessible molecular surface calculated with GEPOL87.

	Probe Radius			
Generation	0.0	1.4	2.8	4.2
1	454.59	1370.42	2635.61	4345.96
2	1510.21	4132.68	7064.97	10496.14
3	4619.03	10970.85	16329.28	22156.25
4	13972.33	27176.48	36462.87	46104.86
	Probe Radius			
Generation	5.6	7.0	8.4	9.8
1	6585.22	9462.27	12848.35	17076.22
2	14649.72	19540.23	25210.93	32019.72
3	28755.81	36258.23	44812.57	54086.09
4	56725.52	68472.88	80972.71	94761.37

Table 15. Regression fits of probe radius versus (accessible surface volume)$^{\frac{1}{3}}$ for the polyether dendrimers.

Generation	B_0^a	B_1^a	R_{Ideal}^b	$R_{B_1}^c$	Reg[d]	GEPOL[e]	Excess[f]
1	9.128	1.706	5.66	5.35	1527.	1370.	0.12
	(0.099)	(0.015)					
2	14.336	1.787	8.89	8.02	4774.	4133.	0.16
	(0.140)	(0.021)					
3	20.571	1.776	12.76	11.58	12260	10970	0.12
	(0.166)	(0.025)					
4	28.368	1.772	17.60	16.01	29360	27180	0.08
	(0.170)	(0.025)					

[a] The value given in parenthesis represents the standard error.

[b] Calculated using ideal slope of 1.612.

[c] Calculated using fit slope, B_1.

[d] Calculated from the regression fit with probe = 1.4 Å.

[e] Calculated using GEPOL87[13] with probe = 1.4 Å.

[f] Fraction of volume that is available to solvent.

Table 16. Volume (Å^3) for the β-alanine dendrimers' accessible molecular surface calculated with GEPOL87.

	Probe Radius			
Generation	0.0	1.4	2.8	4.2
1	282.63	994.04	2145.37	3766.20
2	824.05	2751.68	5611.57	9145.33
3	1928.65	6364.17	12706.59	20255.46
4	4096.89	13192.30	25225.92	38623.18
5	8412.50	26252.35	48339.59	70924.95
6	17351.33	53026.72	92106.31	125804.38
7	34972.90	106197.38	177653.69	233070.69
	Probe Radius			
Generation	5.6	7.0	8.4	9.8
1	5906.48	8665.09	12108.64	16293.96
2	13374.97	18303.49	24252.87	31030.76
3	28627.54	37862.92	48107.71	59165.50
4	52811.78	67872.30	84057.50	101403.63
5	92261.61	113890.90	136317.63	159344.30
6	155793.61	186056.42	216789.08	248711.38
7	277783.28	320233.16	365147.72	406996.88
	Probe Radius			
Generation	11.2	12.6	14.0	
1	21242.91	26982.12	33686.20	
2	38982.20	47832.92	58053.81	
3	71653.23	85590.51	100811.23	
4	119824.20	140141.56	162130.08	
5	184872.42	212534.92	240697.69	
6	283577.41	318908.97	356308.06	
7	453033.31	501130.94	550634.44	

Table 17. Regression fits of probe radius versus (accessible surface volume)$^{\frac{1}{3}}$ for the β-alanine dendrimers.

Generation	B_0^a	B_1^a	R_{Ideal}^b	$R_{B_1}^c$	Reg^d	GEPOL^e	Excess^f
4	27.313	1.954	16.94	13.98	27130	13190	1.06
	(0.298)	(0.028)					
5	34.911	1.961	21.66	17.80	53400	26250	1.03
	(0.263)	(0.024)					
6	43.469	1.971	26.97	22.05	98790	53030	0.86
	(0.286)	(0.027)					
7	55.172	1.923	34.23	28.69	193700	87550	1.21
	(0.302)	(0.028)					

[a] The value given in parenthesis represents the standard error.

[b] Calculated using ideal slope of 1.612.

[c] Calculated using fit slope, B_1.

[d] Calculated from the regression fit with probe = 1.4 Å.

[e] Calculated using GEPOL87[13] with probe = 1.4 Å.

[f] Fraction of volume that is available to solvent.

value of 1.612 expected from the model. The molecular radii determined from the intercept values are in very good agreement with the SEC data and short compared to $D_{end-end}$ and the D_{SA} from surface area analyses.

These linear regression equations may be used in conjunction with the calculated volumes (at p=1.4 Å) to predict the amount of volume not occupied by polymeric material. This analysis suggests that ~50% of the volume contained within these higher generation β-alanine starburst dendrimers is available for solvent or guest molecules.

3.5.2 Normalized Accessible Surface Volumes

The accessible surface volume data (presented in Tables 18 and 19) also allow for the calculation of the change in volume per added monomer. Figures 24 and 25 illustrate this change in normalized V_{AS} plotted as a function of increasing probe size. There should be an increase in volume as probe radius grows. This increase would exhibit, under ideal, noncongested circumstances, a dependence upon the radius, and hence p, to the third power. For the polyether dendrimers, the van der Waals volume (probe = 0.0 Å) is approximately constant for all four cases. As the probe sphere increases, however, only generations 1 and 2 are of the form to permit the volume to show substantial increase with the probe sphere. Generations 3 and 4 both exhibit enough congestion so that the volume per monomer can rise only slightly with an increase in probe.

For the β-alanine series, the change in V_{AS} per monomer does not differ by generation at low values of p. It is only at $p \geq 5.6$ Å that a separation is achieved in the probe vs. ΔV_{AS} plots. As with the ether polymers, it is the lower generation materials (one through four) that structurally permit a substantial increase in normalized volume at high probe radii. Because of the spheroidal

Table 18. Volume ($Å^3$) per added monomer for the polyether dendrimers' accessible molecular surface calculated with GEPOL87.

	Probe Radius			
Transition	0.0	1.4	2.8	4.2
Core → 1	86.39	241.09	422.96	640.58
1 → 2	87.97	230.19	369.11	512.52
2 → 3	86.36	189.95	257.34	323.89
3 → 4	86.60	150.05	186.42	221.75
	Probe Radius			
Transition	5.6	7.0	8.4	9.8
Core → 1	898.01	1195.69	1498.19	1843.74
1 → 2	672.04	839.83	1030.22	1245.29
2 → 3	391.84	464.39	544.49	612.95
3 → 4	258.98	298.28	334.82	376.62

Table 19. Volume (Å^3) per added monomer for the β-alanine dendrimers' accessible molecular surface calculated with GEPOL87.

	Probe Radius			
Transition	0.0	1.4	2.8	4.2
Core → 1	85.85	283.13	572.55	942.34
1 → 2	90.24	292.94	577.70	896.52
2 → 3	92.05	301.04	591.25	925.84
3 → 4	90.34	284.51	521.64	765.32
4 → 5	89.91	272.08	481.53	672.95
5 → 6	93.11	278.90	455.90	571.66
6 → 7	91.78	276.93	445.56	558.68
	Probe Radius			
Transition	5.6	7.0	8.4	9.8
Core → 1	1384.89	1905.04	2525.36	3204.93
1 → 2	1244.75	1606.40	2024.04	2456.13
2 → 3	1271.05	1629.95	1987.90	2344.56
3 → 4	1022.68	1250.39	1497.91	1759.92
4 → 5	821.87	958.72	1088.75	1207.10
5 → 6	661.79	751.72	838.24	930.91
6 → 7	635.36	698.84	772.70	824.79
	Probe Radius			
Transition	11.2	12.6	14.0	
Core → 1	3958.25	4761.27	5632.80	
1 → 2	2956.55	3475.13	4061.27	
2 → 3	2722.59	3146.47	3563.12	
3 → 4	2007.12	2272.96	2554.95	
4 → 5	1355.17	1508.20	1636.83	
5 → 6	1028.18	1108.06	1204.27	
6 → 7	882.58	949.07	1012.12	

Figure 24. The normalized volume versus accessible surface probe radius for the polyether dendrimers. Dashed lines connect the calculated values.

Figure 25. The normalized volume versus accessible surface probe radius for the PAMAM dendrimers. Dashed lines connect the calculated values.

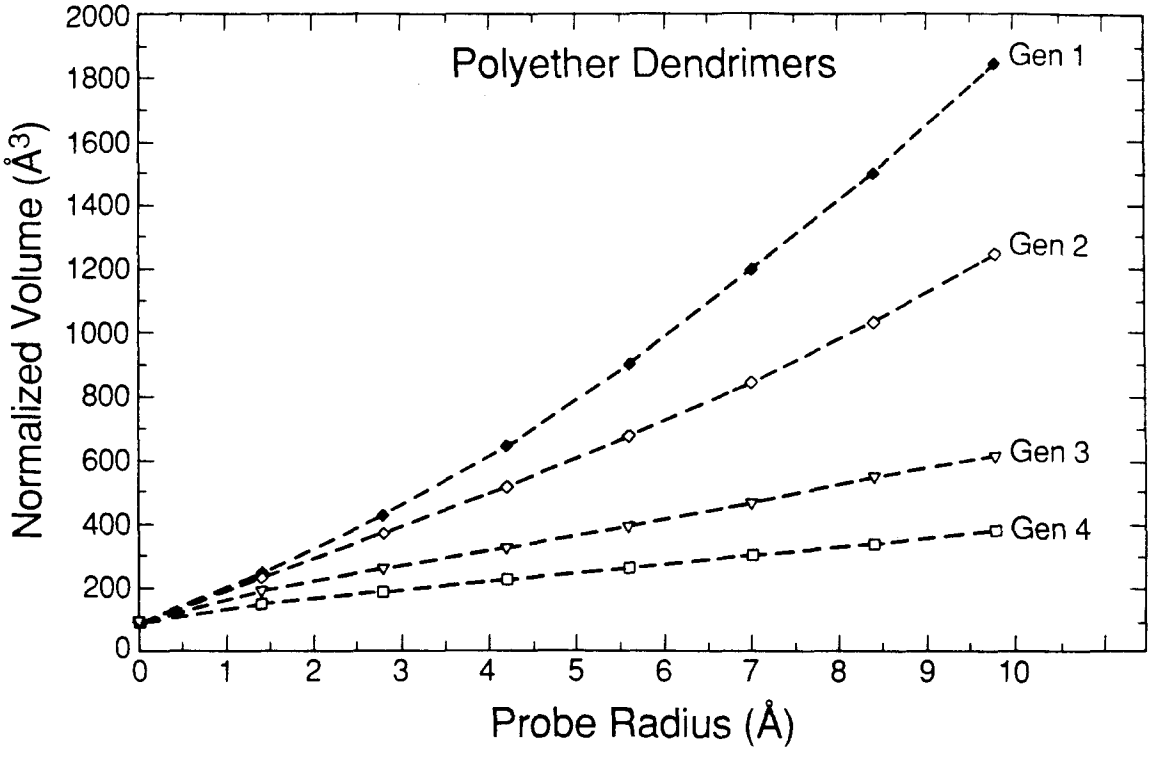


Figure 24.

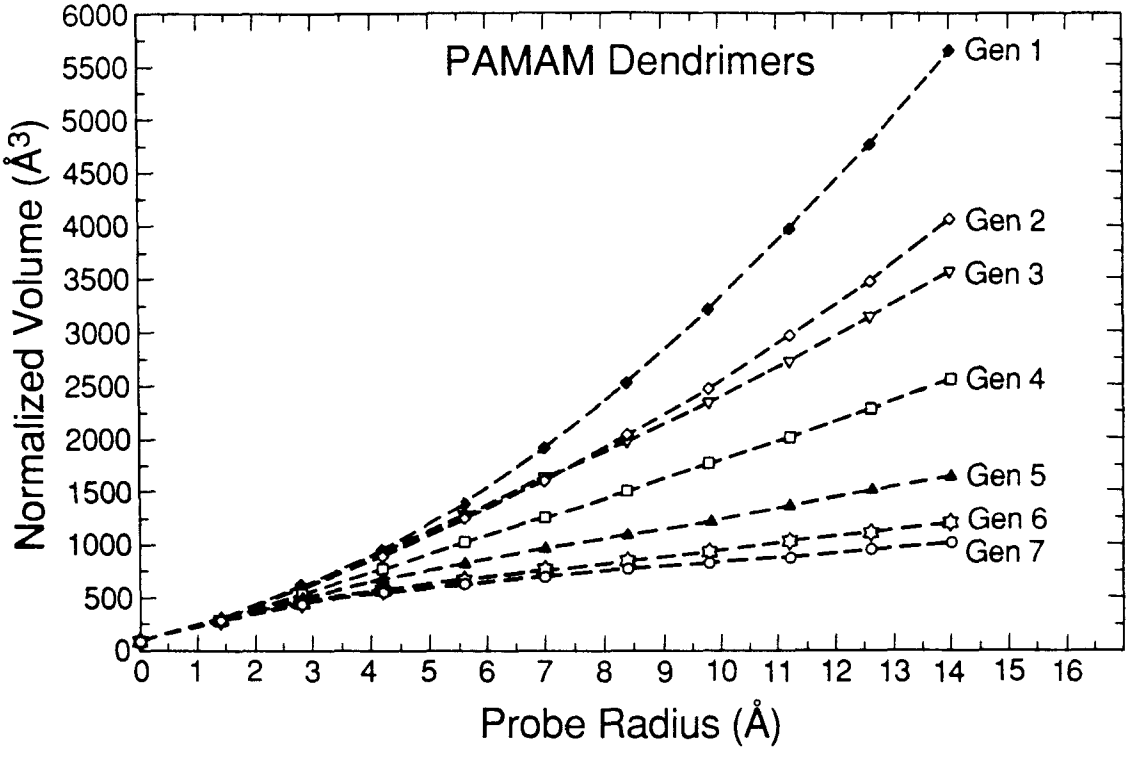


Figure 25.

shape of generations 5 through 7, only modest increases in this volume index are observed.

3.6 Surface vs. Volume Analysis

The analysis of the accessible surface volume, expressed as a function of surface area, also sheds insight on the spherical nature of these macromolecules. For an ideal sphere of radius R, the volume of the accessible surface at probe radius p is given by:

$$V_{AS} = \frac{4}{3}\pi(R + p)^3, \tag{6}$$

while the area mapped out by that molecular surface is:

$$S_{AS} = 4\pi(R + p)^2. \tag{3}$$

Rearranging these two equations for constant value of R + p, i.e., a single probe radius, gives the dependence of volume on surface area, namely:

$$\mathrm{Ln}(V_{AS}) = \frac{3}{2}\mathrm{Ln}(S_{AS}) - Ln(3\sqrt{4\pi}). \tag{9}$$

Thus, for a perfect sphere, a plot of $\mathrm{Ln}(S_{AS})$ vs $\mathrm{Ln}(V_{AS})$ would have a slope of 1.5 and an intercept of ~-2.364.

Polyether

This suggests fitting the accessible surfaces generated for the polyether dendrimers to the equation

$$\mathrm{Ln}(V_{AS}) = B_0 + B_1 \cdot \mathrm{Ln}(S_{AS}). \tag{10}$$

leading to the results presented in Table 20 and in Figure 26. As the radius of the probe sphere used to calculate the accessible surface is increased,

Table 20. Regression fit of Ln (surface area) versus Ln (volume) for the polyether dendrimers.

Probe	B_0^a	B_1^a
0.0	-0.519	1.049
	(0.080)	(0.010)
1.4	-1.854	1.357
	(0.518)	(0.066)
2.8	-2.100	1.429
	(0.297)	(0.037)
4.2	-2.102	1.445
	(0.215)	(0.026)
5.6	-2.179	1.462
	(0.203)	(0.025)
7.0	-2.145	1.462
	(0.189)	(0.022)
8.4	-2.152	1.467
	(0.095)	(0.011)
9.8	-2.250	1.480
	(0.173)	(0.020)

[a] The value given in parenthesis represents the standard error.

Figure 26. The Ln (volume) versus Ln (surface area), at various probe radius values, for the polyether starburst dendrimers. The solid line represents the fit derived for the case of an ideal sphere.

Figure 27. The Ln (volume) versus Ln (surface area), at various probe radius values, for the PAMAM starburst dendrimers. The solid line represents the fit derived for the case of an ideal sphere.

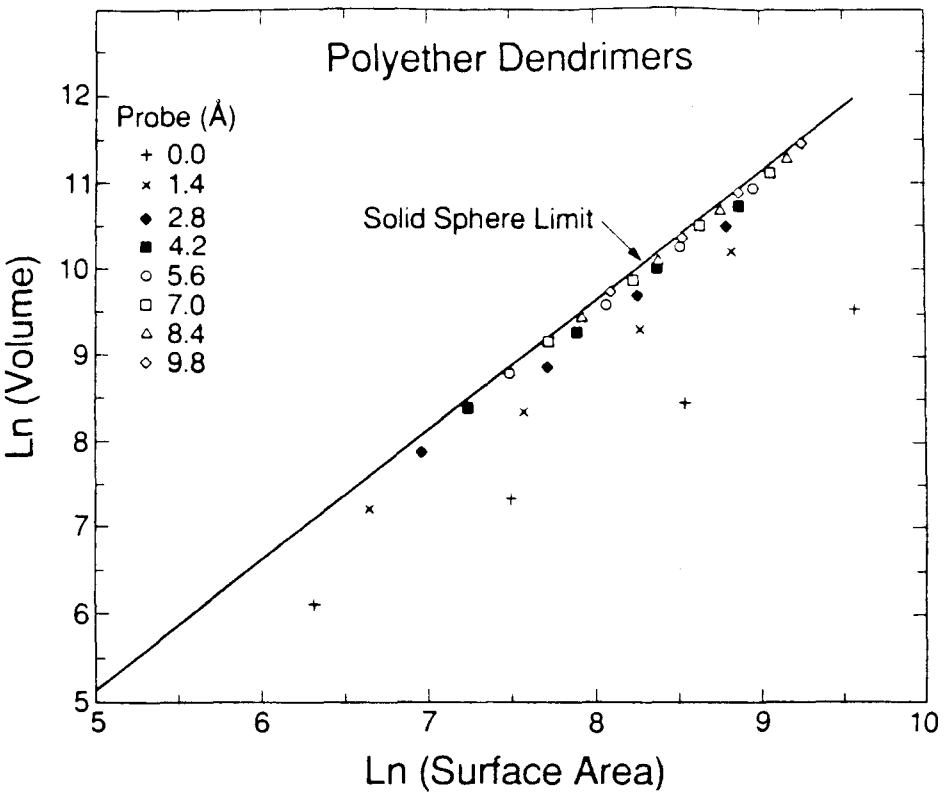


Figure 26.

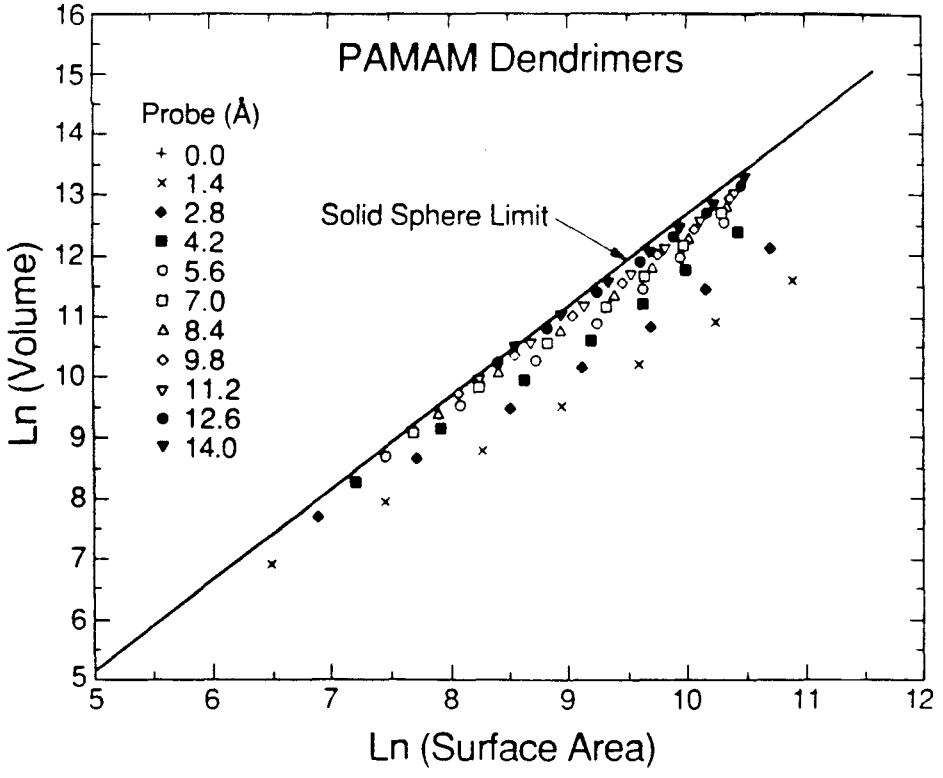


Figure 27.

(1) the value of the intercept increases from -0.519 ($p = 0.0$) to -2.250 ($p = 9.8$) very systematically, approaching the limiting value of -2.364, and

(2) the slope of the line also approaches its limiting value of 1.5, increasing from 1.049 ($p = 0.0$) to 1.480 ($p = 9.8$).

Thus, with increasing size of probe radius, the behavior of the polyether dendrimers approaches that of a perfect sphere.

PAMAM

The situation is a bit more complicated for the β-alanine series. Linear fits of $\mathrm{Ln}(S_{\mathrm{AS}})$ vs. $\mathrm{Ln}(V_{\mathrm{AS}})$ at various p for the *full set* of polyamidoamine polymers produce slopes and intercepts that do approach the spherical limits, although more slowly than those seen with the polyethers (see Tables 20 and 21). Increasing the value of p to 11.2, 12.6 and 14.0 Å verifies this slow convergence towards the ideal values.

If, on the other hand, we partition these polyamino dendrimers into two sets, the early generations (1 through 3) and the later generations (4 through 7); prior to the linear regression studies, our analysis is much different. Under these conditions, we find that the higher generation materials do, in fact, behave more like spherical units. The fitted slopes and intercepts (plotted in Figure 27) now more quickly converge (albeit from the direction opposite that of the polyethers) to the limiting values predicted from the analysis of a sphere. The early generation polymers, however, still exhibit slow convergence of the regression values, indicative of their open, nonspherical nature.

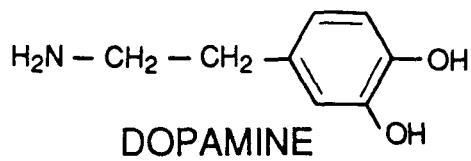
Table 21. Regression fit of Ln (surface area) versus Ln (volume) for the β-alanine dendrimers.

	Full Set		Gen 1 to 3		Gen 4 to 7	
Probe	B_0	B_1	B_0	B_1	B_0	B_1
0.0	-0.262	1.006	-0.332	1.016	-0.146	0.994
	(0.025)	(0.003)	(0.057)	(0.008)	(0.042)	(0.004)
1.4	0.015	1.060	0.120	1.045	-0.080	1.062
	(0.032)	(0.004)	(0.076)	(0.010)	(0.046)	(0.005)
2.8	-0.282	1.149	0.112	1.099	-1.160	1.237
	(0.163)	(0.018)	(0.267)	(0.034)	(0.272)	(0.027)
4.2	-0.911	1.262	-0.267	1.182	-2.780	1.451
	(0.286)	(0.032)	(0.252)	(0.032)	(0.260)	(0.026)
5.6	-1.282	1.328	-0.588	1.244	-3.513	1.555
	(0.331)	(0.036)	(0.268)	(0.033)	(0.312)	(0.032)
7.0	-1.427	1.359	-0.838	1.289	-3.439	1.562
	(0.308)	(0.034)	(0.161)	(0.019)	(0.292)	(0.030)
8.4	-1.509	1.377	-1.073	1.327	-3.169	1.544
	(0.253)	(0.027)	(0.227)	(0.027)	(0.200)	(0.020)
9.8	-1.572	1.391	-1.158	1.344	-2.820	1.516
	(0.207)	(0.221)	(0.116)	(0.014)	(0.357)	(0.036)
11.2	-1.632	1.402	-1.345	1.371	-2.853	1.524
	(0.190)	(0.020)	(0.132)	(0.015)	(0.138)	(0.014)
12.6	-1.650	1.409	-1.400	1.381	-2.449	1.488
	(0.135)	(0.014)	(0.000)	(0.000)	(0.171)	(0.017)
14.0	-1.765	1.423	-1.483	1.392	-2.484	1.494
	(0.138)	(0.014)	(0.000)	(0.000)	(0.328)	(0.032)

[a] The value given in parenthesis represents the standard error.

4. Utilization of β-Alanine Internal Void Spaces

As an illustration of the dimensions of the channels in the β-alanine dendrimers and of how these materials might be used to sequester small molecules, molecular simulation capabilities were used to predict the optimum conformations for several dopamine (**5**) molecules inside this starburst polymeric matrix.



DOPAMINE

5

Dopamine was chosen because it is a good candidate for effective sequesterization in a β-alanine type dendrimer. Its size and shape are suitable for the cavities and channels found in the higher generation β-alanine systems. From a chemical standpoint, dopamine possesses both the hydrogen-bond donor and acceptor sites needed for favorable binding interactions to the polymer's carbonyl, amide, and amino substituents.

Pharmacologically, the ability (i) to selectively deliver dopamine to peripheral kidney receptors without eliciting complications because of the presence of dopamine receptors in the central nervous system (CNS) and (ii) to maintain a supply of nonmetabolized dopamine might be advantageous in cardiovascular hypertension therapy.[18] The additional issues presented by this type of application include the dimensions of the dendrimer-encapsulated dopamine and the targeting of the unit to the appropriate dopamine receptor sites.

Based upon the calculated internal volume of these polymers and the volume of dopamine itself, we estimate that a generation 6 polymer should be capable of holding 20-25 molecules of dopamine. Molecular simulations were carried out

with ten dopamine guest molecules sequestered in the channels of a generation 6 β-alanine host dendrimer. Initially, the polymer matrix was maintained as a rigid entity, while the small molecules optimized their positions within the interior regions. Subsequent calculations then allowed *both* polymeric host and dopamine guest moieties to relax and equilibrate. Figure 28 illustrates this dopamine/dendrimer complex with examples of the conformations adopted by the sequestered molecules and the polymeric material. The vdw surfaces for the components of this complex are included in the photo to highlight the nice fit of dopamine within the dendrimer channels.

After 1.5 and 3.0 ps of simulation time, the environment of the heteroatom units of the ten encapsulated dopamine molecules, i.e., the terminal amino group and the two aromatic hydroxyl substituents, were examined. Of the twenty total dopamine amino group interactions, twelve were involved in hydrogen bonding to polymeric carbonyl oxygen atoms, three were near the dendrimer amide units, and five occupied sites distant enough from the polymer to be accessible to the solvent. For the forty possible hydroxyl group interactions, twenty-two were capable of donating a hydrogen for interaction with either amino nitrogens (15 times) or carbonyl oxygens (7 times), and fourteen were close enough to participate in hydrogen bonding by acting as acceptor to amide protons. In only four instances were the hydroxyl groups free from polymer interaction to hydrogen bond with solvent.

The dimensions of the higher generation β-alanine dendrimers (see Figure 29 with the minimum diameter plotted as a function of generation) are such that these dopamine/dendrimer complexes would not readily cross the CNS blood-brain barrier.

The advantage of the dendrimer technology is that we can modify the surface

Figure 28. Illustration of dopamine (red) sequestered in the internal channels and pockets of a generation 6 β-alanine starburst dendrimer. The inner generations of the polymer (one through five) are shown in light blue, while the outer, sixth, generation is displayed in green.

Figure 30. A catechol-modified (yellow terminal groups) β-alanine starburst dendrimer (light blue) with bound dopamine molecules (red).

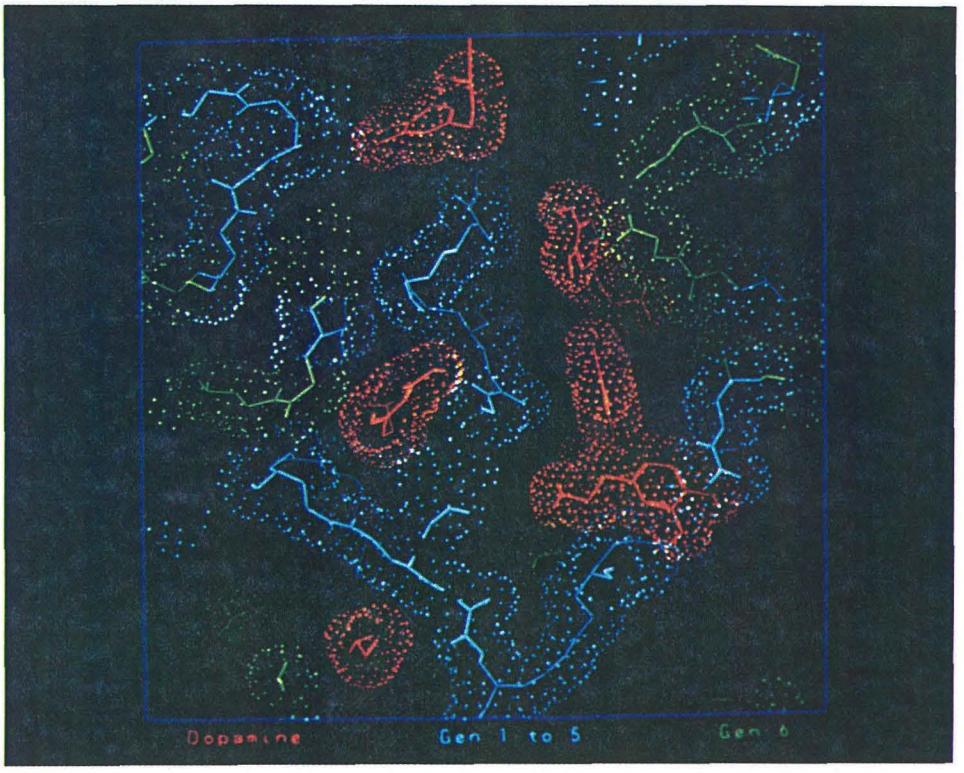


Figure 28.

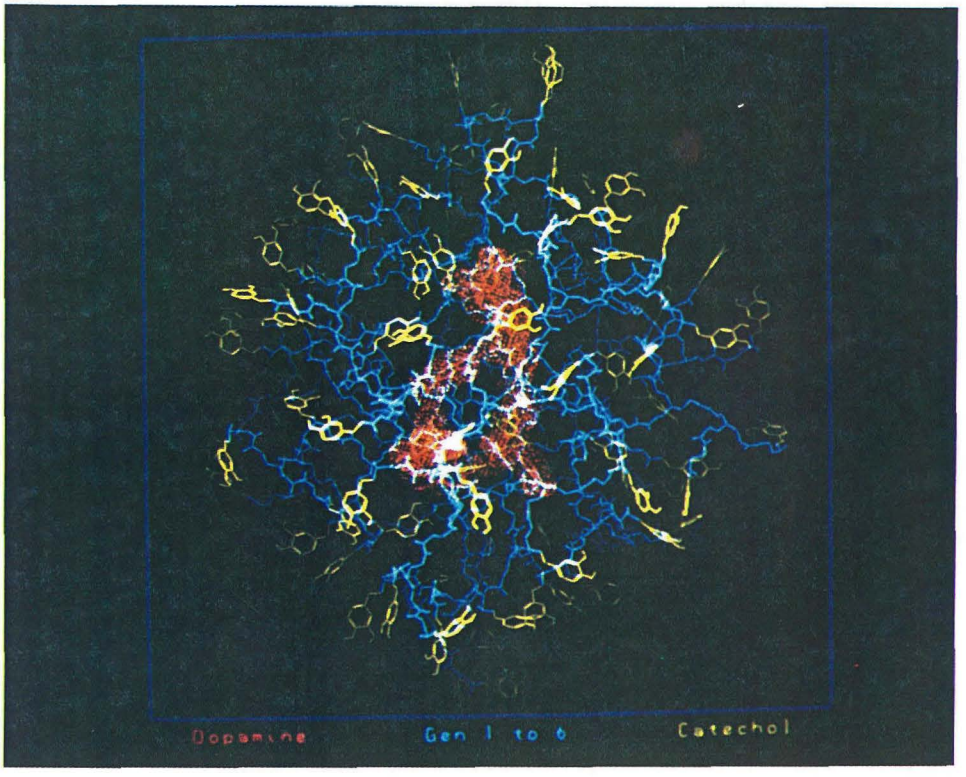


Figure 30.

Figure 29. The minimal diameter versus generation for the β-alanine starburst dendrimers.

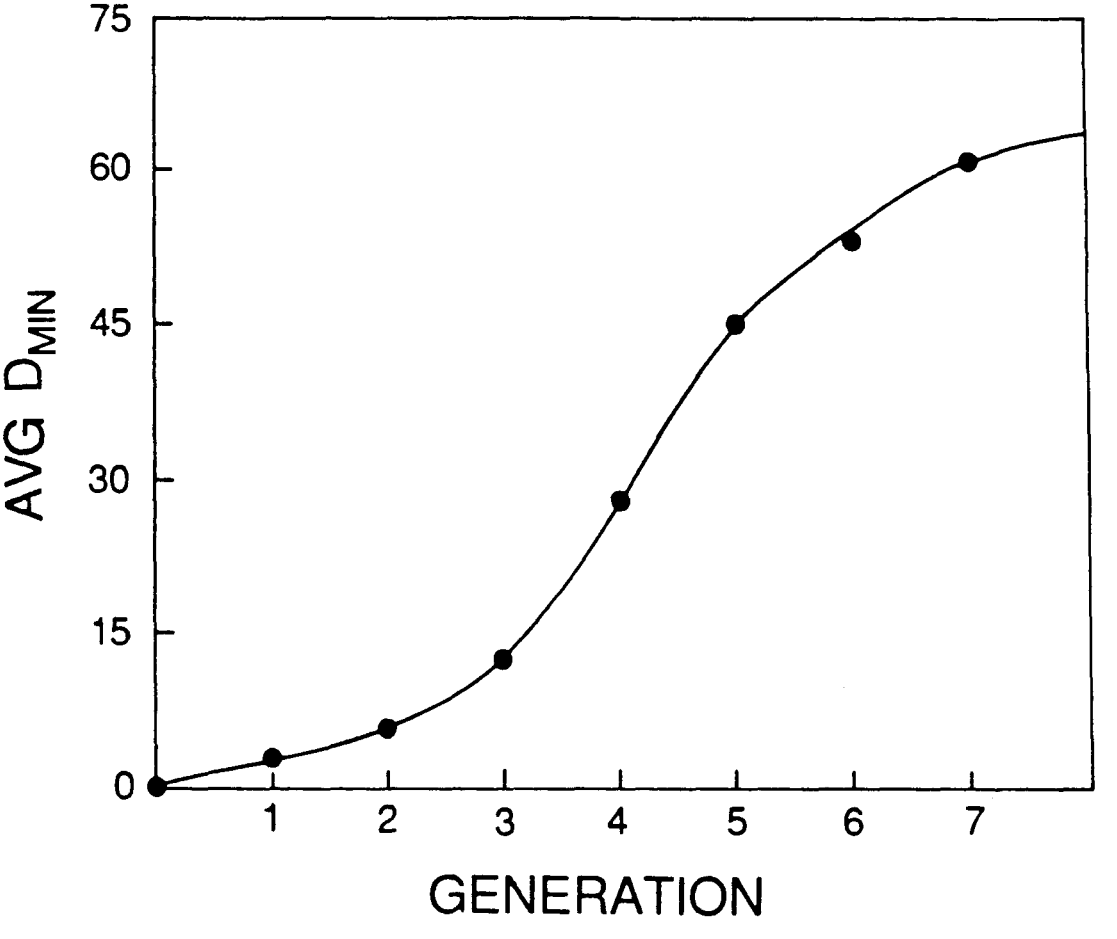
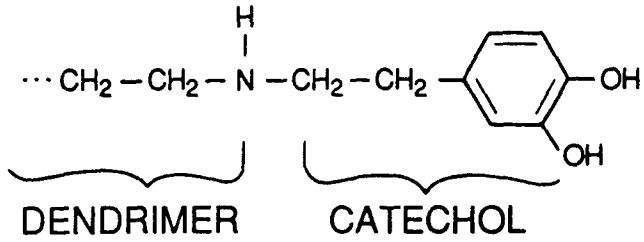


Figure 29.

layer to ensure delivery to the desired site. Thus, condensing dihydroxyphenyl acetaldehyde with the terminal amino group leads effectively to dopaminelike surface catechol fragments, as shown in (**6**).



6

Figure 30 illustrates this hybrid dendrimer complex. The catechol modification to the dendrimer should allow it to recognize dopamine receptor sites and bind sufficiently well to help increase the concentration near these sites. However, the composite dendrimer should be displaced by dopamine as it diffuses out of the dendrimer. Thus, the dendrimer serves as: (1) a polymer encapsulation matrix to protect the catecholamine from rapid metabolic inactivation, (2) a block to transport across the blood-brain barrier, and (3) a concentration enhancer near the receptor sites.

These studies provide structural models that should be useful for analyzing the dopamine/dendrimer systems. The next step is to test the effectiveness of these modified materials for encapsulation of dopamine (and related materials) and to determine how effectively they are delivered to the kidney centers. As such experiments proceed, continuing simulation will be useful in providing a quantitative framework for understanding various results.

The judicious selection of a core unit, internal monomer subunits, and terminal monomer fragments should allow for the design of dendrimers to complex with specific guest molecules and to deliver them to specific sites. It will, of

course, be essential to develop synthetic schemes that allow for chemical control and fidelity.

5. Conclusions

The molecular simulations indicate a qualitative change in the overall shape and character of the β-alanine dendrimers as the polymer grows from generation 1 to generation 7. The shape transition from an open, extended structure to a spheroidal moiety takes place at the fourth generation. Concurrent with this shape change is the introduction of internal cavities and channels in the polymer matrix. These interior regions possess a surface area approximately equivalent to the amount found on the exterior of the polymer spheroid. Similarly, the volume associated with these solvent-filled void spaces is $\sim$50% of that contained within the full spheroid. These internal regions can be used for sequestering small molecules, as illustrated by the example of the dopamine-loaded and catechol-modified delivery dendrimer. Materials of this type could be designed to encapsulate selected target molecules or to contain sites for modification to produce catalytically active buried pockets.

The penta-erythritol polyether dendrimers were found to possess structures and properties quite different from those of the polyamidoamine series. The structures adopted by the polyethers are dense and spherelike early in the polymer growth cascade. Only negligible amounts of interior surface area and volume were found for these materials. Estimates of the amount of surface area per hydroxyl head group indicated that this generation 4 dendrimer should not be structurally stable, consistent with experimental observation.

Notes and References

(1.) D. A. Tomalia, H. Baker, J. Dewald, M. Hall, G. Kallos, S. Martin, J. Roeck, P. Smith. *Polym. J.*, Tokyo, **1985**, *17*, 117.

(2.) D. A. Tomalia, H. Baker, J. Dewald, M. Hall, G. Kallos, S. Martin, J. Roeck, J. Ryder, P. Smith. *Macromolecules* **1986**, *19*, 2466.

(3.) D.A. Tomalia, M. Hall, D.M. Hedstrand. *J. Am. Chem. Soc.* **1987**, *109*, 1601.

(4.) D. A. Tomalia, V. Berry, M. Hall, D. M. Hedstrand. *Macromolecules*, **1987**, *20*, 1164.

(5.) H. Hall, A. Padias, R. McConnell, D. A. Tomalia. *J. Org. Chem.*, **1987**, *52*, 5305.

(6.) P.B. Smith, S. J. Martin, M. J. Hall, D. A. Tomalia. *Applied Polymer Analysis and Characterization.* Edited by J. Mitchell, Jr. Hanser Publishers: New York (1987), p. 357.

(7.) D. A. Tomalia, D. M. Hedstrand, L. R. Wilson. *Kirk-Othmer Encyclopedia of Polymer Science and Engineering*, **1989**, in press.

(8.) L. Wilson and D. Downing, personal communication.

(9.) POLYGRAF is an interactive molecular mechanics/graphics program available from Molecular Simulation, Inc. Pasadena, CA. All simulations were performed on DEC VAX 8650 and Alliant FX8/8 computers, and structures were displayed on Evans & Sutherland PS330 and PS390 series graphics systems.

(10.) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr. P. Weiner. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(11.) F. M. Richards. *Ann. Rev. Biophys. Bioeng.* **1977**, *6*, 151.

(12.) M. L. Connolly. *J. Appl. Cryst.* **1983**, *16*, 548.

(13.) J. L. Pascual-Ahuir, E. Silla. *QCPE Program Number 554.*

(14.) NDIV specifies the division level for the triangles on the surface of the atomic spheres. Accuracy improves as NDIV is increased, but cpu time also increases with NDIV. Generation 1 β-alanine dendrimer was used to investigate NDIV=1,2,3,4,5. The calculated surface areas and volumes were well converged at NDIV=3. With NDIV=4 or 5, any improvement in the calculated values was less than 0.1% but the cpu utilization had increased by a factor of 3 (NDIV=4) or 10 (NDIV=5). Hence, it was determined that the recommended value of NDIV=3 was most suitable.

(15.) A.M. Naylor, W.A. Goddard III, G. E. Keifer, D.A. Tomalia. *J. Am. Chem. Soc.*, **1989**, in press.

(16.) PLOTIT Interactive Graphics and Statistics available from Scientific Programming Enterprises, Haslett MI.

(17.) C. Tanford, *The Hydrophobic Effect. A Formation of Micelles and Biological Membranes.* Wiley Interscience: New York (1980).

(18.) J. W. Kebabian, T. Agui, J. C. van Oene, K. Shigematsu, J. M. Saavedra. *Trends in Pharmacological Sciences* **1986**, 96.

Section III

Dihydrofolate Reductase:

Molecular Simulation of Wild-type and

Mutant Enzyme Complexes

Portions of the text of this section will compose an article coauthored with William A. Goddard III and Stephen J.Benkovic. It is to be submitted to the *Journal of the American Chemical Society.*

Abstract:

Molecular modeling and simulation tools were used to investigate various Dihydrofolate Reductase (DHFR) complexes.

Three distinct sources of DHFR were analyzed in homology studies. The sequence homology between *E. coli* and *L. casei* proteins is less than 30%. Our investigations indicate that the *chemical homology* for the catalytic pocket is ~70%, explaining the remarkable similarity in their steady-state kinetic profiles. The inclusion of *Chicken* DHFR yields only ~60% chemical homology with the bacterial forms, suggesting subtle differences in the character of the active site. This molecular modeling and subsequent simulation were used to engineer *E. coli*·*Chicken* hybrid enzymes targeted to reduce folate.

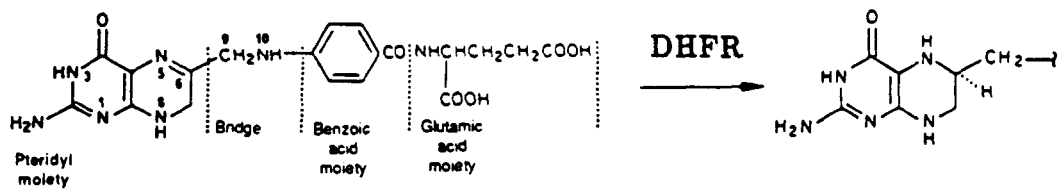
The structure and properties of DHFR complexes (*E. coli*, single-site mutants, and *Chicken* loop *E. coli* hybrids) were studied using molecular dynamics simulations. The calculations on the wild-type *E. coli* enzyme indicate: (a) the energetically favored conformation of bound methotrexate's (MTX) pteridine ring is that found in the crystal studies; (b) the active site region accommodates either the xtl or the anti form of both MTX and dihydrofolate (DHF); and (c) the Anti-DHF ternary complex is favored for catalysis with a ground-state hydride transfer distance of ~3.6 Å.

The investigations on single-site mutants revealed structural changes that explain the experimental kinetic data. Phe-31→Tyr-31 introduces reorientation of substrate binding that is consistent with the observed increase in the product off-rate (the rate-limiting step). Modifications of the alkyl side chain at position 54 (Leu-54→Ile-54 or Leu-54→Gly-54) increase the separation between the co-factor (nicotinamide) and substrate (pteridine) reaction centers, thus increasing the rate of hydride transfer sufficiently so that it becomes rate-limiting.

1. Introduction

Over the 30 years since the discovery that the cancer therapy agent methotrexate (MTX) acts by inhibiting the transformation of dihydrofolate (DHF) to tetrahydrofolate (THF), thereby inhibiting Dihydrofolate Reductase, there have been numerous experiments providing detailed data on many aspects of various forms of the enzyme and its chemistry.[1] This wealth of data makes it an ideal system for our interests, which are to examine how changes in the active site of an enzyme affect function and reactivity.

Dihydrofolate Reductase (DHFR) catalyzes the reduction of 7,8-dihydrofolate (DHF; **1**) to 5,6,7,8-tetrahydrofolate (THF; **2**), as depicted below.



1 **2**

This reduction requires: (1) a hydride (H^-) to be transferred to the substrate from a nicotinamide adenine dinucleotide phosphate (NADPH) cofactor source that is bound in a ternary complex with DHFR and DHF, and (2) a proton (H^+) to be supplied from some source (possibly H_2O). Figure 1 illustrates the DHFR·DHF·NADPH ternary complex. This structure is one of the products of the work described in this document.

The product THF is an important coenzyme in a number of one-carbon transfer reactions. Its most important role, though, is in the synthesis of thymidylate (dTMP) from deoxyuridylate (dUMP) by thymidylate synthetase. In this metabolic process, THF not only is the source of a carbon fragment but is also used as a reductant. Hence, it is used in substrate, rather than coenzyme, quan-

Figure 1. The active site of wild-type *E. coli* Dihydrofolate Reductase with dihydrofolate and NADPH (both shown in red) bound. The solvent surface of the active site is mapped out by the series of dots.

Figure 3. Wild-type *E. coli* DHFR with MTX (red) and NADPH (light blue) bound in their respective sites. Asp-27 (dark blue), involved in protonation of the substrate pteridine ring, is shown salt-bridged to the MTX. Phe-31 (green) and Leu-54 (blue) are strictly conserved residues that form part of the substrate binding site.

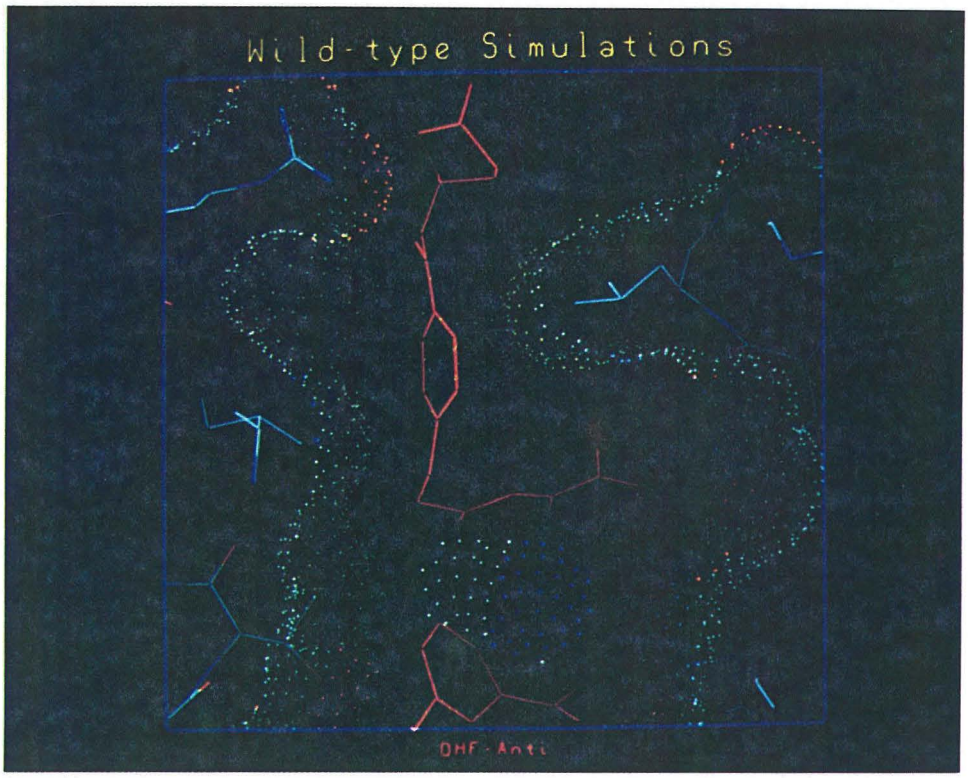


Figure 1.

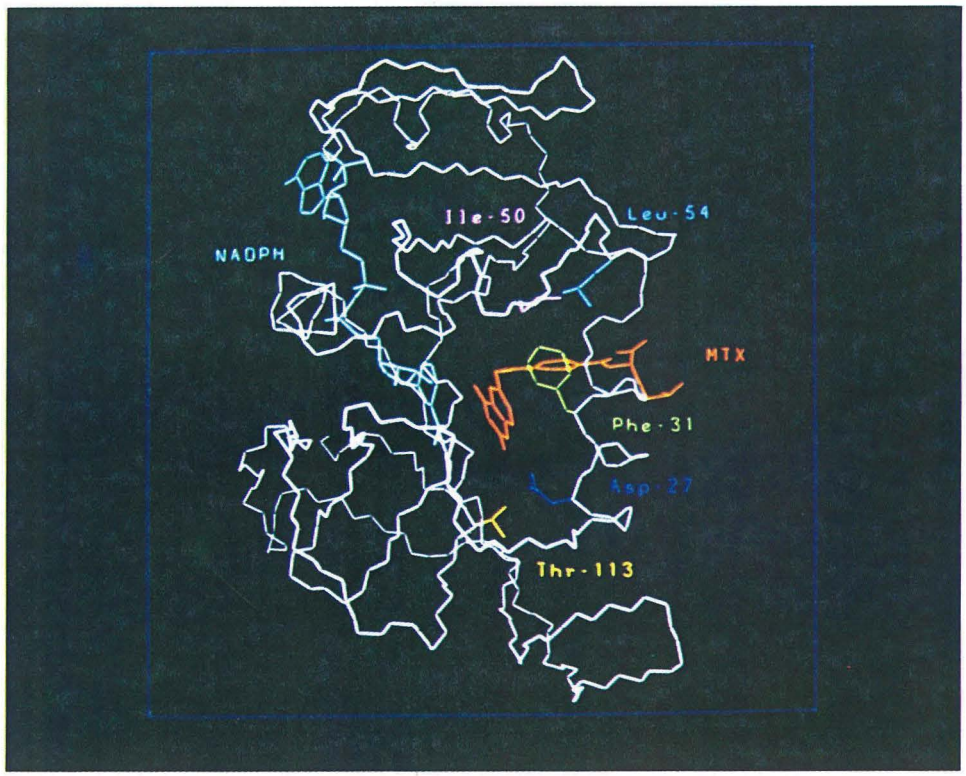
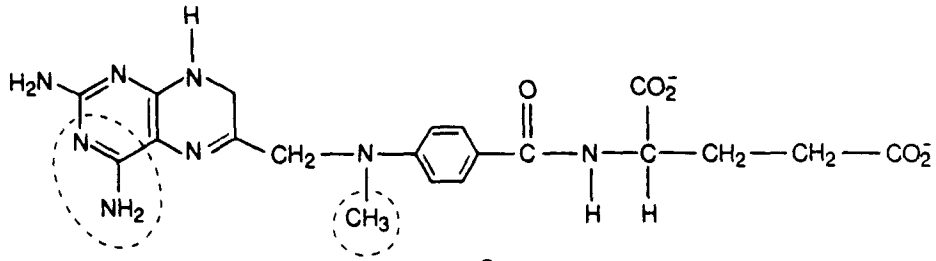


Figure 3.

tities. The THF pool must be maintained by the DHFR-catalyzed reaction. This metabolic cycle is depicted in Figure 2. Inhibition of DHFR leads to a deficiency of dTMP and ultimately to a disruption of nucleic acid biosynthesis. It is this mechanism that is the biochemical basis for the antifolate drugs that have found therapeutic applications.[1]

The known inhibitors of DHFR may be divided into classical and nonclassical subsets. The classical inhibitors are closely related structurally to the folic acid natural substrates of DHFR. Thus, they are expected to bind to the enzyme in a manner quite similar to normal substrates but cannot be easily reduced, preventing the DHFR from transforming DHF to THF. Methotrexate (MTX; **3**), shown below, is one important example of a classical DHFR inhibitor.



3

Many nonclassical inhibitors, chemically distinct from the folic acid derivatives, have been identified.[1,2] These agents may well bind to the enzyme in a manner quite different from that of the natural substrates.

Recent advances in experimental methodologies have made possible the investigation of structure-function relationships in this enzyme system. The crystal structure of DHFR from two bacterial sources (*E. coli* and *L. casei*) has been solved, refined to a 1.7Å resolution,[3-6] and made available to the scientific community through the Brookhaven Protein Database. The study of DHFR from *L. casei* involved the ternary complex of enzyme, MTX (inhibitor), and NADPH (cofactor). The crystal structure of *E. coli* DHFR was determined as a dimer of

Figure 2. Metabolic cycle encompassing the reduction of DHF to THF by DHFR.

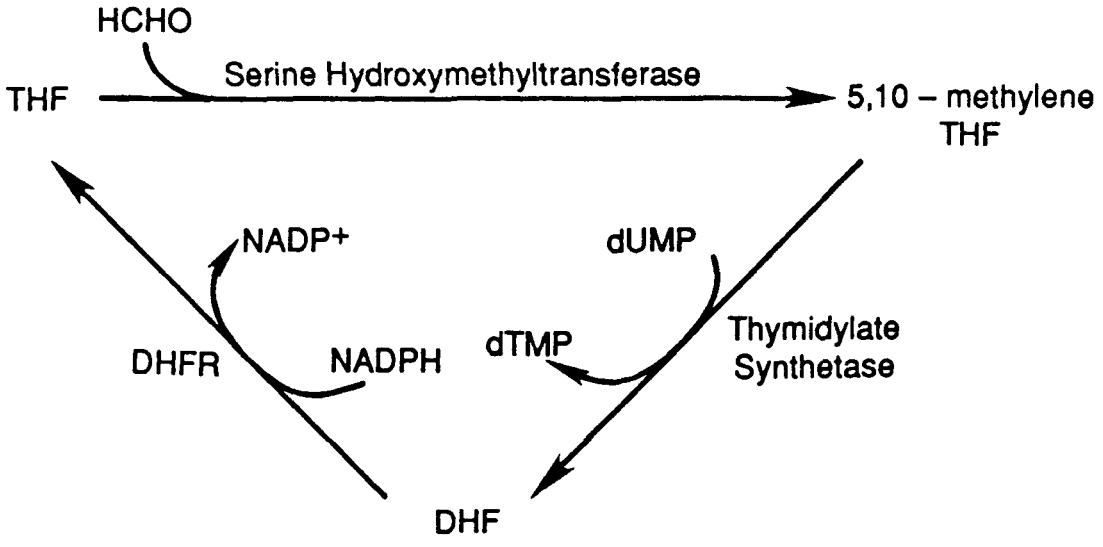

HCHO
Serine Hydroxymethyltransferase
THF
5,10 – methylene
THF
NADP+
dUMP
DHFR
NADPH
dTMP
Thymidylate
Synthetase
DHF

Figure 2.

the one-to-one complex of enzyme with MTX.

The gene for *E. coli* DHFR production has been located and sequenced,[7] thus allowing use of site-directed mutagenesis to synthesize mutants for directly assessing the roles of specific amino acids in catalysis, folding, and stability of the enzyme protein.[8-15] The full steady-state kinetic profile of *E. coli* DHFR has been recently refined and reported.[11]

As a result, DHFR provides an excellent system for a collaborative experimental/theoretical program designed to probe the details of DHFR's catalytic requirements and properties. The experimental component of the project is centered at The Pennsylvania State University under the direction of Professor Stephen J. Benkovic. The theoretical portion of the research project is being done within the Molecular Simulation Facility at the California Institute of Technology under the supervision of William A. Goddard III. The work presented here focuses on three aspects of this collaborative effort.

First, we present the analysis of the sequence, structure, and chemical homology of three forms of DHFR from different species (*E. coli* and *L. casei* bacterial sequences and the *Chicken* form). Experimental kinetic studies indicate a remarkable similarity in the free-energy profile for the reduction of dihydrofolate by *E. coli* and *L. casei* DHFR, despite a mere 27% sequence homology (*vide infra*). Computer graphics and macromolecular docking, combined with homology assignments, offer an explanation for such congruence in the catalytic reactivity of the bacterial systems. The analysis is then extended to include the *Chicken* form of the protein, which has a distinctly different reaction profile and is more similar in function to mammalian DHFR. This interspecies comparison provides the opportunity to begin piecing together an understanding of what contributes to the subtle differences in reductive chemistry exhibited by these enzymes.

The second subject involves molecular simulation studies of wild-type *E. coli* DHFR·MTX binary and DHFR·DHF·NADPH ternary complexes. These investigations offer insights on the catalytic reduction of dihydrofolate as well as controls for computer mutagenesis experiments. A brief discussion of the simulation methods employed in the theoretical investigations of various DHFR complexes is also included.

And finally, the third segment focuses on the selected modification of the *E. coli* DHFR, using both single-site mutation and protein-loop substitutions. The site-directed mutagenesis studies were undertaken to probe the role(s) of specific amino acid residues in the functioning of ***E. coli*** DHFR. Figure 3 illustrates the location within the protein, and proximity to the substrate binding site, of several of these targeted amino acids. Segment mutants were constructed to engineer a bacterial DHFR protein with properties more similar to those of the mammalian enzyme (which has the ability to reduce folate to dihydrofolate). These hybrid proteins are based on the bacterial DHFR (from ***E. coli***) where a loop, with amino acid sequence identical to the corresponding region of the ***Chicken*** DHFR, has been inserted into the protein matrix.

Before embarking on discussions of the modeling and simulation work, it is important to put the relevant experimental studies into perspective. The next section presents a synopsis of the DHFR experimental information germane to this work.

2. Pertinent Experimental DHFR Studies

The abundant experimental data on DHFR provide both inspiration and detailed checks upon theoretical studies. The crystal structures of DHFR from two bacterial sources have been resolved, reported, and coordinates made available through the Brookhaven Data Base.[3-6] Isotope studies have identified the stereochemistry of the tetrahydrofolate product, providing information on the orientation of bound, reactive substrate.[16] The kinetic profiles of the enzyme have been determined under various conditions, providing mechanistic details.[11,13,17]

2.1 Crystal Structures

From the crystal studies, it is apparent that the backbone structures of both the *E. coli* and *L. casei* sources of DHFR are quite similar despite an only ~29% direct sequence homology between the two forms. Figure 4 illustrates this overall structural similarity in the folding pattern of these bacterial DHFR proteins. The backbone is composed of ~35% β-sheet structure with the remainder consisting of four α-helices and interconnecting loops.

The conformation of the bound methotrexate inhibitor is also quite similar in the two crystals. The 2,4-diaminopteridine ring sits in a pocket, which allows for specific hydrogen bond and electrostatic interactions of an Asp residue (27 in *E. coli* = *Ec*; 26 in *L. casei* = *Lc*) with the amino substituent at position 2 and with N1. A conserved Thr residue (113 *Ec*, 116 *Ec*) and an intervening fixed water molecule are also involved in hydrogen-bonds to N2. The pyrazine portion of MTX is not directly hydrogen-bonded to the protein but is involved in solvent H-bonding. A strictly conserved Phe ring (31 *Ec*, 30 *Lc*) is involved in nonpolar contacts with both the pterin and benzoyl rings of the bound inhibitor. The glutamate side chain of the MTX has different conformations in the two

Figure 4. A. Protein backbone structure for *E. coli* DHFR·MTX binary crystal complex. The MTX inhibitor is located in the center of the figure.

Figure 4. B. Protein backbone structure for *L. casei* DHFR·MTX·NADPH ternary crystal complex. The MTX inhibitor and NADPH cofactor are located in the middle of the figure.

Figure 4. C. Protein backbone structure for *Chicken* DHFR·NADPH binary crystal complex. The NADPH cofactor is found at the left center of the figure.

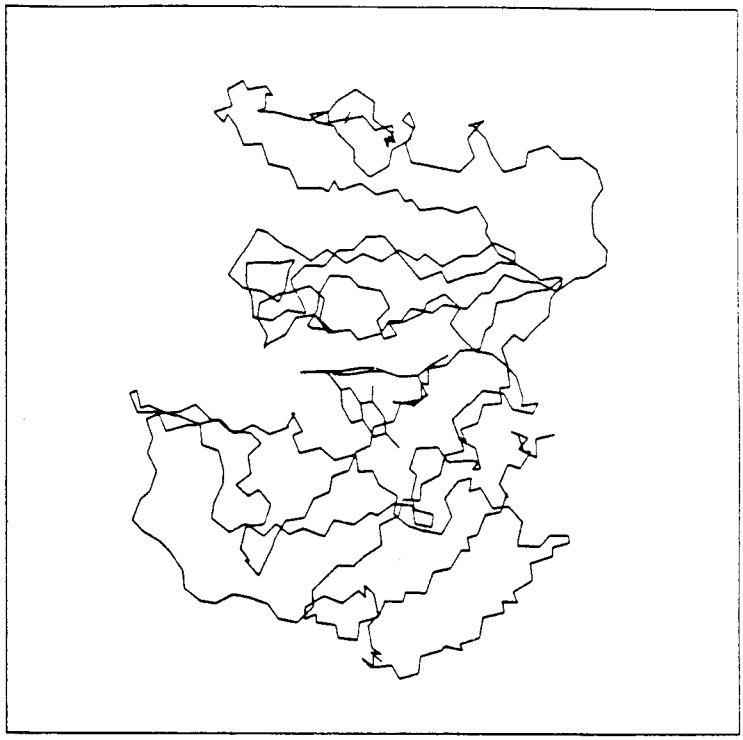


Figure 4. A.

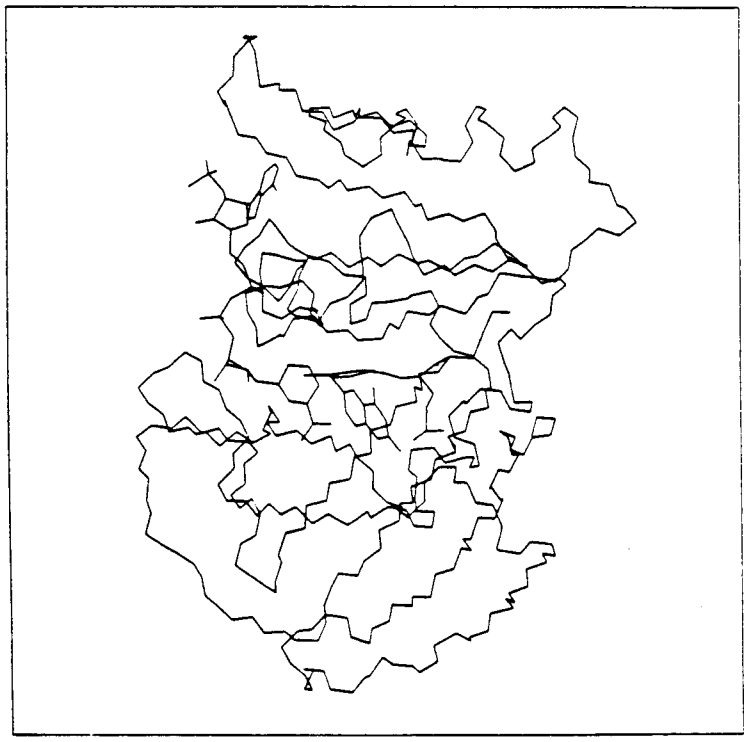


Figure 4. B.

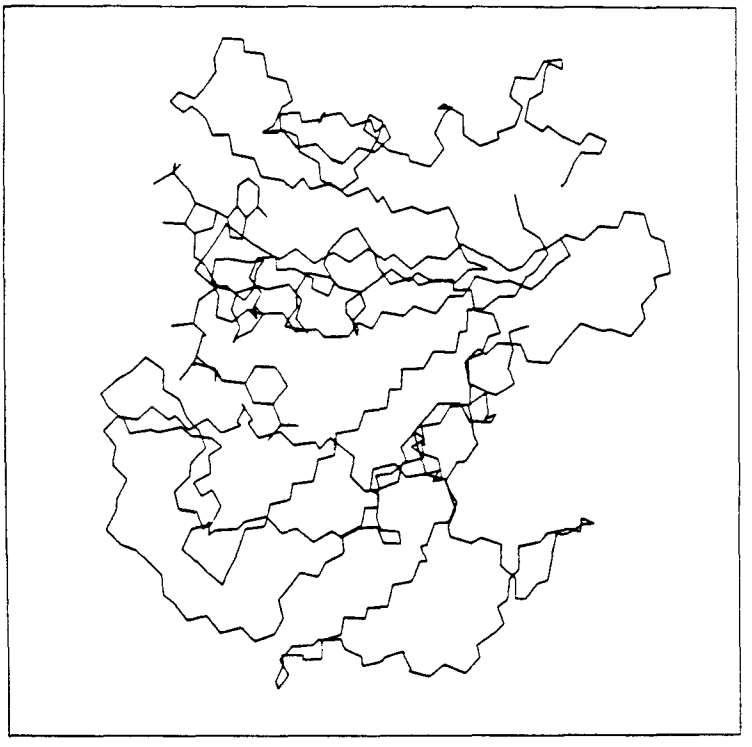


Figure 4. C.

complexes, even though it binds, in both instances, to the conserved Arg-57 through the α carboxylate group.

An NADPH binding site distinct from the substrate binding site does not exist in the DHFR system. Instead there are substrate and cofactor domains that contain overlapping portions of the primary sequence, as illustrated by the display in Figure 3. No major reorganization of secondary structure is noted between the *L casei* (with NADPH bound) and *E. coli* (lacking the NADPH cofactor) forms. From the crystal study of the ternary complex, the only contacts found between the adenine ring of NADPH and the enzyme are hydrophobic. The pyrophosphate bridge region of NADPH is stabilized by hydrogen-bonding to enzyme backbone amido groups, bridging solvent molecules, and the side chains of Arg-44 (*Lc*) and Thr-45 (*Lc*). The binding of the nicotinamide ribose appears to be fairly specific, making hydrophobic contacts with a number of invariant residues (Ile-3, Gly-14, and Gly-99 *Lc*) and H-bonding through its hydroxyl groups to the carbonyl oxygen of His-18 (*Lc*) and the side chains of Asp-125 and Ser-48. The adenosine ribose interactions with DHFR appear to be weaker and less specific, containing only one hydrogen bond between the protein and the ribose, mediated by a water molecule. The 2′-phosphate group is involved in hydrogen bonding and/or salt bridges to Arg-43, Thr-63, His-64, and Gln-65 (all *Lc*). The nicotinamide functional group of the NADPH resides in a pocket created by predominantly invariant amino acid residues. The A side of the ring is in contact with the pteridine binding site, while the B side is surrounded by a group of hydrophobic side chains that exclude solvent molecules. An interesting characteristic of this binding site is that three protein oxygen atoms are positioned in close proximity to C2, C4 and C6 of the nicotinamide ring. The carboxamide group is coplanar with the nicatinamide ring but is rotated 180 degrees from the

minimal energy solution conformation.

In addition to this published crystallographic information on the bacterial DHFRs, unpublished structural data for this enzyme from the *Chicken* source were generously made available from Matthews and coworkers in prerelease form.[18] These experimental data for the *Chicken* protein were obtained for the DHFR · NADPH binary complex. The coordinates for the structure were refined to 1.7 Å resolution. The overall fold of the *Chicken* form matches quite well with that of the bacterial forms, as illustrated in Figure 4. The major differences between the bacterial and *Chicken* tertiary structures involve inserted amino acid residues that are typically confined to loop regions on the exterior portion of the protein. The differences in sequence between the *Chicken* and bacterial DHFRs, as well as the variations introduced by the loop insertions, will be elaborated upon further below.

2.2 Isotope Studies

Young and coworkers[16] used deuterium isotope studies, in conjunction with absorption and nuclear magnetic resonance spectroscopies, to identify the stereochemistry of 5,6,7,8-tetrahydrofolate produced by Dihydrofolate Reductase.

The reduction of DHF using NADPH and DHFR produces THF that gives the biologically active S diasteriomer when converted into folinic acid. This defines the absolute stereochemistry at C6 of the enzymatic DHF as S. Therefore, the *re*-face of DHF is the side subject to attack by the NADPH cofactor, as diagrammed in Figure 5. It is also known that the 4-*pro-R* hydrogen of the nicontinamide ring is the one involved in transfer to DHF, thus establishing that the orientation adopted by the two reacting rings during catalysis is different from the orientation of MTX's bound pteridine moiety.

Figure 5. A schematic view of the orientation of the nicotinamide (NADPH) and pteridine (DHF) reacting centers involved in hydride transfer.

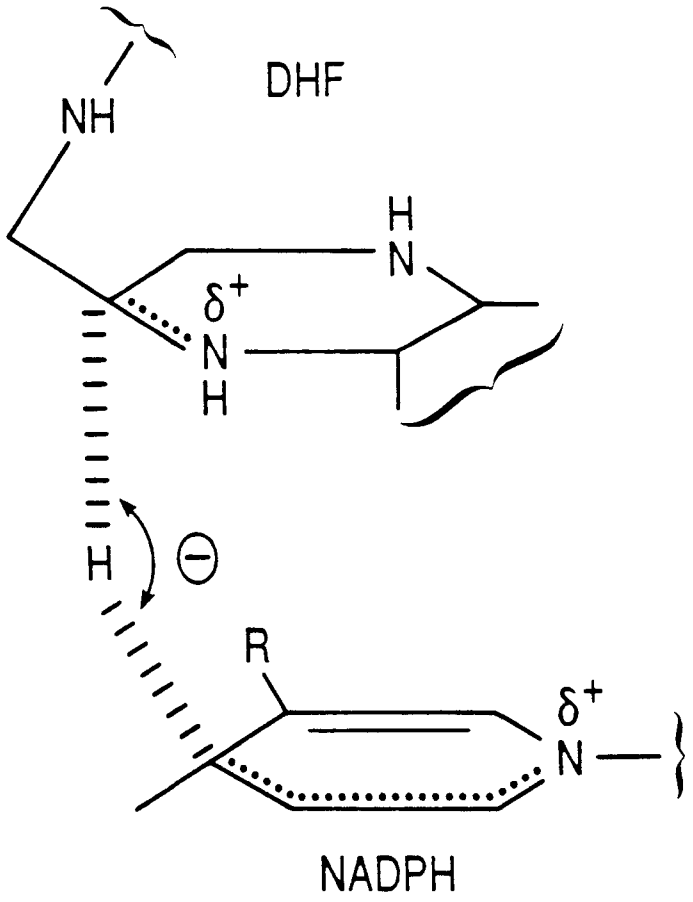
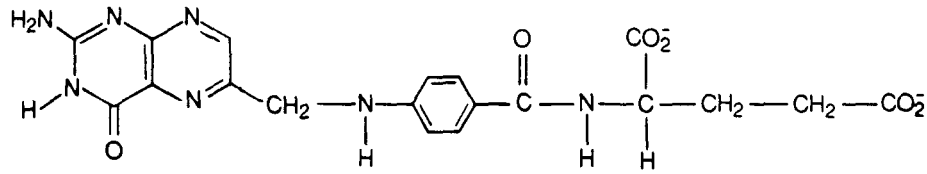


Figure 5.

DHFR from certain sources will also reduce folate (4) to DHF.[19]



4

With the *L. casei* form, this reduction is ~100 times slower than that of DHF. During the enzymatic reduction of folate, it is again the 4-*pro-R* hydrogen transferred to the pteridine ring. Spectroscopy reveals that the isotope is transferred to the *si* face of the C7 in folate. The enzyme catalyzed hydride transfer, for the reduction of both DHF and folate, involves the same face of NADPH and the same face of the folic acid substrate. Thus, there is no major difference in the orientation of *reactive* DHF or folate substrate binding.

2.3 Kinetic Profile

A profile of the steady-state kinetics of the wild-type *E. coli* form of DHFR was recently reported by Fierke, Johnson, and Benkovic.[11] This study used stopped-flow fluorescence and absorbence spectroscopies in combination with competetive binding assays to measure the rates of ligand association and dissociation as well as the binding constants of key intermediates in the catalytic sequence. A diagram of this steady-state profile of the native form of *E. coli* DHFR is shown in Figure 6.

A number of surprising findings came of this work. The Benkovic studies[11] revealed that the catalytic step, i.e., the N5-C6 double bond reduction, proceeds at a steady-state rate of 950 sec^{-1}. The steady-state rate of THF *product release*, however, is only 12.5 sec^{-1}, thus making product release the rate-limiting step!

Figure 6. Steady-state kinetic scheme for wild-type *E. coli* DHFR.

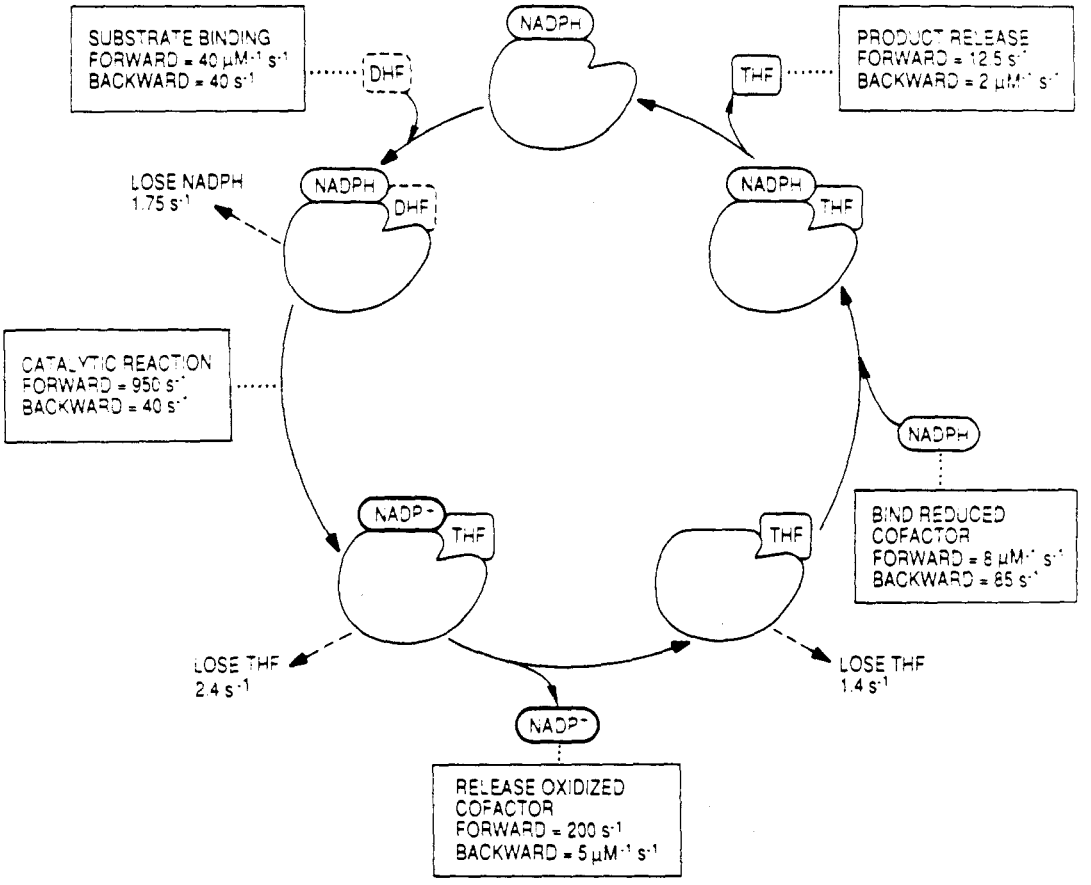

SUBSTRATE BINDING
FORWARD = 40 μM⁻¹ s⁻¹
BACKWARD = 40 s⁻¹
DHF
NADPH
PRODUCT RELEASE
FORWARD = 12.5 s⁻¹
BACKWARD = 2 μM⁻¹ s⁻¹
THF
LOSE NADPH
1.75 s⁻¹
NADPH
DHF
NADPH
THF
CATALYTIC REACTION
FORWARD = 950 s⁻¹
BACKWARD = 40 s⁻¹
NADPH
BIND REDUCED
COFACTOR
FORWARD = 8 μM⁻¹ s⁻¹
BACKWARD = 85 s⁻¹
NADP⁺
THF
THF
LOSE THF
2.4 s⁻¹
LOSE THF
1.4 s⁻¹
NADP⁺
RELEASE OXIDIZED
COFACTOR
FORWARD = 200 s⁻¹
BACKWARD = 5 μM⁻¹ s⁻¹

Figure 6.

The rate of THF release from the enzyme depends significantly upon whether the cofactor binding site is occupied. The off-rate for THF from the DHFR · THF · $NADP^+$ ternary complex is 2.4 sec^{-1}, while the loss of $NADP^+$ decreases this rate to 1.4 sec^{-1}. With the addition of NADPH to the enzyme complex, the product off-rate is markedly increased to 12.5 sec^{-1}. Thus, THF release is facilitated by the presence of the *reduced* NADPH cofactor, indicative of a second role for the NADPH cofactor.

Mutant forms of *E. coli* DHFR have been studied in a series of experiments quite similar to those used to determine the wild-type, steady-state profile.[8,9,11-15] The results germane to our simulation studies are presented in Table 1. These values indicate that modifications at position 54 (a strictly conserved leucine residue) to Ile or Gly lead to the following changes in the kinetics: (a) the substrate binding affinity decreases by more than one order of magnitude; (b) the hydride transfer rate decreases by more than 30 fold; and (c) the steady-state rate of product release increases by more than eight fold. For both single-site mutants, the rate-limiting step becomes hydride transfer, rather than product release, thus decoupling the transformation rate from the process of cofactor binding.

2.4 The Active Site Asp-27

From chemical precedent, the mechanism for imine reduction favors preprotonation of the pteridine N5 followed by hydride transfer. Examination of the pH dependence of the steady-state turnover showed that the rate constant for the isolated hydride transfer step is dependent on an acidic group with $pK_a \approx 6.5$, a value that is the same for free, binary, or ternary complex forms of the enzyme. The identity of this acidic active site residue was clearly shown to be Asp-27 (*E.*

Table 1. Steady-state kinetic data[9,14] for wild-type and position 54 mutants of *E. coli* DHFR.

	Leu-54	Gly-54	Ile-54
	Binding (μM)		
E_{DHF}	0.21	1.9	350
	Dissociation (sec^{-1})		
E_{DHF}	22	$\gg$300	300
E_{THF}	1.4	>300	60
$E_{THF}^{NADP^+}$	2.5	-	100
E_{THF}^{NADPH}	12.3	-	90
	Hydride Transfer (sec^{-1})		
	950	29	31

coli) by its replacement with Asn or Ser, resulting in two mutant DHFRs that require preprotonated DHF for catalysis.[10]

3. Homology Studies

Steady-state kinetic studies of the different DHFR system from *E. coli* and *L. casei* (only 27% sequence homology) reveal a remarkable similarity in the free-energy profiles for DHF to THF.[13] In order to understand why this congruence in catalytic profile occurs for such different proteins, we undertook investigations of the similarity in sequence, structure, and chemical homology of the *E. coli* and *L. casei* forms of DHFR. Studies using computer graphics with macromolecular docking were combined with both sequence and chemical homology assignments to offer an explanation for the agreement in catalytic reactivity of these bacterial systems. This work is described in detail below and presented as portions of the manuscripts included in Appendices III and IV.

This structure-sequence study was also extended to include the *Chicken* DHFR form which, along with other mammalian DHFR systems, has a reaction profile quite distinct from the bacterial systems. Such an interspecies comparison provides the opportunity to begin piecing together an understanding of the origins of the subtle differences in reductive chemistry exhibited by these enzymes.

3.1 *E. coli* and *L. casei* Bacterial Forms

The molecular display and manipulation facilities of BIOGRAF[20] were used to analyze these bacterial DHFR crystal complexes. This analysis involved docking the enzyme complexes, determining the substrate and cofactor binding sites, and comparing the amino acid residues and active site molecular surfaces.

3.1.1 Coordinate Transformation

The initial step was to dock the crystallographic form of the *entire L. casei* ternary complex onto the coordinate space of the *E. coli* binary complex so as to match the MTX of the two systems. The results of this docking exercise are

illustrated in Figures 7 and 8. Here the *E. coli* backbone structure is shown in yellow with the bound MTX in red, while the docked *L. casei* segment is drawn in light blue. Figure 7 clearly illustrates the structural agreement in overall folding of these two bacterial forms of DHFR. Figure 8 highlights the similarity of MTX binding and the criterion used for docking the *L. casei* complex to that of the *E. coli*.

3.1.2 Identification of the Site Regions

In order to focus on the sequence homology of the active site region, we defined the ***substrate active site*** as all residues having any atom within 7 Å of any atom of the MTX. This resulting substrate active site has 52 amino acids (out of a total of 159 for the *E. coli* or 162 for the *L. casei*) as listed in Table 2.

The ***NADPH active site*** was identified in an analogous manner but based solely on the position of the *L. casei*-bound NADPH. For these two DHFR complexes, a slight discrepancy in the definition of the NADPH active site was encountered. The differences were due to the fact that the *L. casei* form was indeed a ternary complex with NADPH bound, while the *E. coli* structure lacked this small molecule. The *L. casei* system had the protein segment from Asp-16 through Trp-21 securing the position of the NADPH moiety, while the binary *E. coli* protein permitted the analogous Glu-17 through Trp-22 region to be extended and somewhat more removed from the empty cofactor site. Hence, the 7Å distance cutoff produced the full inclusion of *L. casei* Asp-16 to Trp-21 while missing a few of the residues in the *E. coli* complex. To insure appropriate analyses, the full *E. coli* segment from Glu-17 through Trp-22 was added to those amino acids identified to be within 7 Å, thus producing an NADPH active site containing 59 amino acid residues. These residues are listed in Table 3.

Figure 7. Comparison of the overall protein fold of the *E. coli* (yellow) and *L. casei* (light blue) forms of DHFR. The MTX inhibitor (red from *E. coli* and blue from *L. casei*) and NADPH cofactor (blue from *L.casei*) are shown in their cavities.

Figure 8. Active site comparison of the *E. coli* and *L. casei* bacterial forms of DHFR. The docking criterion is clearly illustrated by the agreement in structure of the two bound MTX molecules (red from *E. coli* and blue from *L. casei*). The solvent surfaces are also displayed highlighting the high degree of agreement between the two active sites.

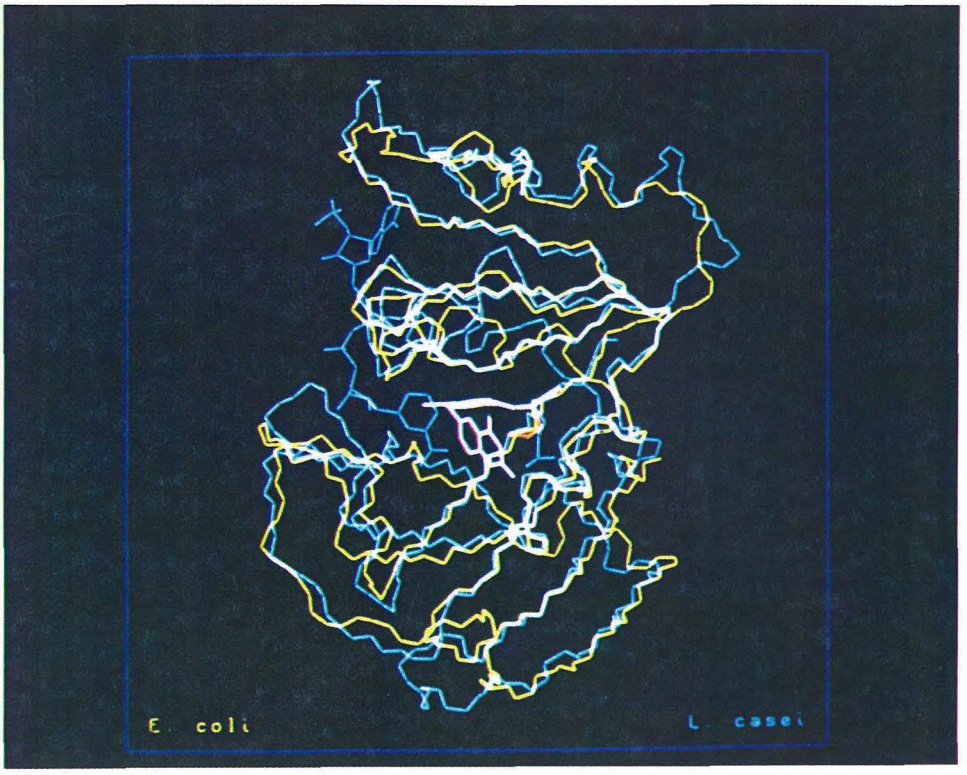


Figure 7.

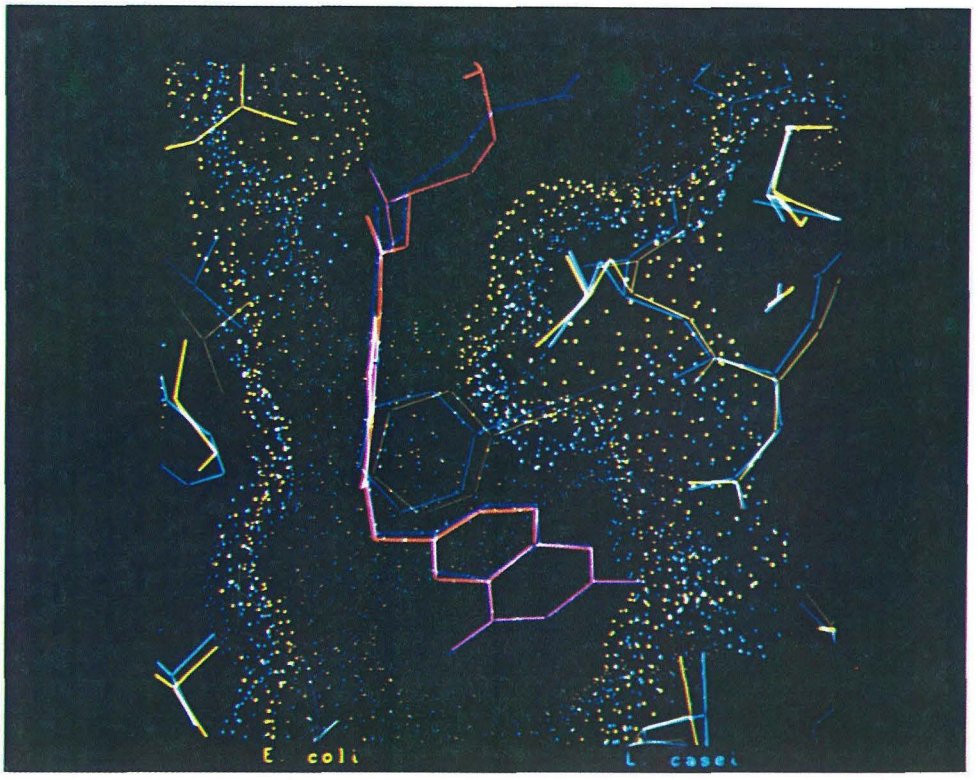


Figure 8.

Table 2. Homology of the substrate binding site[a] for DHFR.

E. coli	*L. casei*	Class[a]	Involvement in the Active Site
Leu-4	Phe-3	Back	deep in pterin pocket
Ile-5	Leu-4	Whole	side chains allign near diamino portion of site and Phe-31(30)
Ala-6	Trp-5	Back	βC allignment; Trp ring away from pterin site, stacked with Phe-103
Ala-7	Ala-6	Exact	methyl group part of pterin pocket
Leu-8	Gln-7	Back	interacts with other residues forming pocket
Ala-9	Asn-8	Back	sides removed from pterin site
Asn-18[b]	Gly-17	–	{ interacts with
Ala-19[b]	His-18	–	Nicatinamide ribose }
Met-20[b]	Leu-19	Whole	side chain into the pterin/nic/benz site
Pro-21[b]	Pro-20	Exact	hugs NADPH ribose
Trp-22	Trp-21	Exact	edge of pterin pocket
Asn-23[b]	His-22	Whole	under the pterin/benz site
Leu-24	Leu-23	Exact	pterin pocket
Pro-25	Pro-24	Exact	base of helix
Ala-26	Asp-25	Back	on the side of helix away from site
Asp-27	Asp-26	Exact	Bridge to pteridine
Leu-28	Leu-27	Exact	MTX kink region
Ala-29	His-28	Sides	imidazole interacting with δ-CO2-
Trp-30	Tyr-29	Whole	rings allign; edge of pterin site
Phe-31	Phe-30	Exact	pteridine hydrophobic pocket
Lys-32	Arg-31	Whole	straddles both CO2- of MTX

Table 2. (continued)

E. coli	*L. casei*	Class[a]	Involvement in the Active Site
Arg-33	Ala-32	Back	on helix side away from site
Asn-34	Gln-33	Back	on helix side away from site
Thr-35	Thr-34	Exact	side toward MTX; buffered by Phe-30/31
Leu-36	Val-35	Whole	edge of glutamate site
Asp-37	Gly-36	Back	Asp side chain at surface
Val-40	Met-39	Whole	interacting with Phe-31/30
Met-42	Val-41	Whole	removed from site region
Thr-46	Thr-45	Exact	nic site near benz/amine portion
Trp-47	Tyr-46	Back	behind MTX benz site
Glu-48	Glu-47	Exact	toward NADPH site
Ser-49	Ser-48	Exact	toward NADPH site
Ile-50	Phe-49	Whole	behind benz of MTX
Gly-51	Pro-50	Whole	kink; edge of benz site
	Lys-51	-	
Arg-52	Arg-52	Exact	*E. coli* with δ-CO2- of MTX; *L. casei* pointing in different direction
Pro-53	Pro-53	Exact	kink away from site
Leu-54	Leu-54	Exact	benz/glutamate site
Pro-55	Pro-55	Exact	end of glutamate site
Gly-56	Glu-56	Back	Glu side chain out on surface of enzyme
Arg-57	Arg-57	Exact	salt bridge to MTX CO2- (CT)
Met-92	Val-95	Whole	far back side of pterin site pocket
Val-93	Ile-96	Back	into bulk of protein
Ile-94	Ala-97	Whole	hydrophobic pterin pocket
Gly-95	Gly-98	Exact	{backbone forms side to
Gly-96	Gly-99	Exact	nic/pterin(N4-N5) site}

Table 2. (continued)

E. coli	*L. casei*	Class[a]	Involvement in the Active Site
Tyr-100	Phe-103	Whole	point toward pterin/nic site
Tyr-111	Leu-114	Back	{ good backbone and side allignment;
Leu-112	Val-115	Back	composes the β sheet below Leu-4
Thr-113	Thr-116	Exact	to Leu-8; near/behind 2-4 diamino site }
His-114	Arg-117	Back	on the outside of the enzyme away from the active site
Ile-115	Leu-118	Whole	deep in pterin site near MTX N1 (near Trp-22/21)
Phe-153	Tyr-155	Sides	next to Asp-26/27

[a] As discussed in Section 3.1.3, classification scheme defined as:

Exact: the two amino acid residues are identical in type.

Back: the backbone atoms align, while the side chain atoms are oriented *away* from the active site.

Sides: the side chains align toward the bound substrate.

Whole: the backbone atoms and side chain atoms align as best possible with the side chain oriented *toward* the active site.

[b] Not well aligned because of the absence of NADPH in the *E. coli* crystal structure.

Table 3. Homology of the cofactor binding site[a] for DHFR.

E. coli	*L. casei*	Class[a]	Involvement in the Active Site
Ile-5	Leu-4	Whole	away from Nic near Pterin site
Ala-6	Trp-5	Back	*E. coli* Ala-6 + Leu-8 side = *L. casei* Trp-5 side + Gln-7 side
Ala-7	Ala-6	Exact	near amide of Nic ring
Leu-8	Gln-7	Back	*E. coli* Ala-6 + Leu-8 side = *L. casei* Trp-5 side + Gln-7 side
Ala-9	Asn-8	Back	Far removed from site
Val-10	Arg-9	Back	Far removed from site
Val-13	Leu-12	Back	Pterin pocket
Ile-14	Ile-13	Exact	floor of Nic ring site
Gly-15	Gly-14	Exact	edge of Nic site
Met-16	Lys-15	Back	sides both point to outside of enzyme
Glu-17[b]	Asp-16	Back	side chains out to surface of enzyme
Asn-18[b]	Gly-17	Back	Nicotinamide ribose site
Ala-19[b]	His-18	Back	side chains at enzyme surface
Met-20[b]	Leu-19	Whole	Nicotinamide/pterin site
Pro-21[b]	Pro-20	Exact	Edge of nicotinamide pocket
Trp-22[b]	Trp-21	Exact	Nicotinamide/pterin pocket
Asn-23	His-22	Whole	MTX glutamate site
Leu-24	Leu-23	Exact	Pterin site
Asp-27	Asp-26	Exact	toward pterin ring area
Phe-31	Phe-30	Exact	pterin site

Table 3. (continued)

E. coli	*L. casei*	Class[a]	Involvement in the Active Site
Met-42	Val-41	Whole	located behind Nic ribose
Gly-43	Gly-42	Exact	part of adenosine binding site
Arg-44	Arg-43	Exact	adenosine phosphate neutralizer
His-45	Arg-44	Whole	diphosphate site charge neutralizer
Thr-46	Thr-45	Exact	back of diphos/Nic site
Trp-47	Tyr-46	Back	sides stack with Phe-49/Ile-50
Glu-48	Glu-47	Exact	sides point out to surface of enzyme; 2nd segment of amino acids over from diphos site
Ser-49	Ser-48	Exact	OH points toward Nic ribose
Ile-50	Phe-49	Whole	back part of Nicotinmide/benz site
Ile-61	Val-61	Whole	far removed from Aden ribose site
Leu-62	Leu-62	Exact	Adenosine binding pocket
Ser-63	Thr-63	Whole	-OH interacts with Aden phos
Ser-64	His-64	Whole	His ring stacks with Aden base, Ser -OH in same place
Gln-65	Gln-65	Exact	displacement by the Adenosine phosphate group
Pro-66	Glu-66	Back	out to surface away from site
Gly-67	Asp-67	Back	out to surface away from site
Thr-68	Tyr-68	Sides	*E. coli* Thr-68 + Trp-47 sides = *L. casei* Tyr-68 + Tyr-46 sides
Trp-74	Val-75	Sides	*E. coli* Trp-74 + Thr-68 sides = *L. casei* Val-75 + Tyr-68 sides
Val-75	Val-76	Exact	sides into bulk of enzyme
Lys-76	His-77	Back	sides away from site cavity at surface of enzyme
Ser-77	Asp-78	Back	side chains away from site at enzyme surface
Val-78	Val-79	Exact	edge of Adenenosine base site
Asp-79	Ala-80	Back	CO2- to outside of enzyme; away from site

Table 3. (continued)

E. coli	*L. casei*	Class[a]	Involvement in the Active Site
Val-93	Ile-96	Back	into bulk of protein
Ile-94	Ala-97	Whole	*E. coli* Ile-94 + Ile-50 sides = *L. casei* Phe-49 + Ala-97 sides; edge of Pterin/Nic site
Gly-95	Gly-98	Exact	near Nic/Pterin interface
Gly-96	Gly-99	Exact	base of diphosphate binding area
Gly-97	Ala-100	Back	CH3 points away from diphosphate site
Arg-98	Gln-101	Whole	interact with either diphosphate or Adenosine phosphate
Val-99	Ile-102	Whole	form part of Adenenosine base and ribose site
Tyr-100	Phe-103	Whole	nicotinamide site
Glu-101	Thr-104	Back	sides away from site area
Gln-102	Ala-105	Whole	might be some adjustment when NADPH binds
Phe-103	Phe-106	Exact	buried into bulk of enzyme
Ile-115	Leu-118	Sides	Nic end behind Trp22/21
Asp-122	Asp-125	Exact	H-bonding to nicotinamide ribose
Thr-123	Thr-126	Exact	behind Nic ribose and diphos regions
His-124	Lys-127	Back	sides to surface and far removed from site
Phe-125	Met-128	Whole	far, far back of Nic site

[a] As discussed in Section 3.1.3, classification scheme defined as:

Exact: the two amino acid residues are identical in type.

Back: the backbone atoms align, while the side chain atoms are oriented *away* from the active site.

Sides: the side chains align toward the bound substrate.

Whole: the backbone atoms and side chain atoms align as best possible with the side chain oriented *toward* the active site.

[b] Not well aligned because of the absence of NADPH in the *E. coli* crystal structure.

3.1.3 Comparison of the Site Regions

In comparing the amino acids for different DHFRs, we considered corresponding residues, as those having the same positional and functional role on the basis of type, alignment, and role at the surface of the binding sites. Tables 2 and 3 include this comparison information for the methotrexate/dihydrofolate and NADPH sites, respectively. For the MTX/DHF site, 20 of the 52 amino acids are identical, leading to a 38% sequence homology. For the NADPH site, 22 of the 59 residues are the same, leading to a sequence homology of 37%. Combining the sequence data for the MTX/DHF and NADPH active sites produces a *catalytic pocket* composed of 85 amino acid residues, 31 of which are conserved between the *E. coli* and the *L. casei* forms (36% sequence homology).

To facilitate and refine this active site comparison, a new classification scheme was devised for comparison of residues between *E. coli* and *L. casei*. If the corresponding residues are exactly the same for both forms, e.g., Ala-6 *Ec* and Ala-7 *Lc*, the residue match is considered as "*exact*," as in the normal definition of homology. If, on the other hand, the two residue types differ, we define three possible classification types:

(1) *whole*: where the backbone *and* the side chain atoms are well aligned between the two forms and where the side chain is oriented *toward* the active site;

(2) *side*: where the side chains align toward the bound substrate and/or cofactor but without reasonable agreement of the backbone sites;

(3) *back:* where the backbone atoms are in very good positional agreement, while the side chain fragments are oriented *away from* the active site into the bulk of the enzyme or at the surface of the protein;

(4) *none:* neither backbone or side chain is aligned.

The *chemical homology* scale is defined as consisting of all residues that

are either *exact* (if identical) or of type *back* (if different). Using the chemical homology scale, an additional 14 residues in the MTX/DHF pocket and 19 amino acids in the NADPH binding site (see Tables 2 and 3) are homologous. Thus, the *chemical homology* for the *E. coli* and *L. casei* forms is: (a) 65% for the substrate binding region, (b) 69% for the cofactor site, and (c) 69% for the entire catalytic pocket, even though the sequence homology is less than 30%.

This close chemical homology may be visualized with the solvent-accessible molecular surface[21,22] of the substrate and cofactor pockets. These surfaces were calculated with a 1.4 Å probe sphere, consistent with a water molecule probe, and displayed with BIOGRAF. The areas mapped out by these surfaces agree within 95%. The agreement between the molecular surfaces for the *E. coli* and the *L. casei* forms is illustrated in Figure 8.

3.2 Chicken DHFR

Because of the importance of mammalian DHFR in cancer therapy, and because of the dramatic differences in the reaction profile for mammalian and bacterial DHFR, it is quite important to understand the subtle differences between these forms of DHFR. Such knowledge might facilitate designing therapeutic agents that would be species-selective for interaction with DHFR. For this reason, this structure-sequence analysis was extended to include a *Chicken* DHFR, the only nonbacterial form of DHFR for which there is detailed structural information.

The crystal structure of the DHFR·NADPH binary complex has been solved to a 1.7 Å resolution by Matthews and coworkers[18] and the coordinates generously made available to our lab. The overall sequence homology for this protein with the bacterial forms is <30%. As with the two bacterial forms of this en-

zyme, however, the fold of the *Chicken* form of the protein has a close similarity to that of both the *E. coli* and *L. casei* structures (see Figure 4).

The macromolecular docking of the *Chicken* binary complex was carried out to optimize the fit of the NADPH of *Chicken* with that of *L. casei*. The cofactor-binding site of chicken DHFR was defined as described above, using *Chicken* NADPH. To define the MTX substrate-binding cavities of *Chicken* DHFR, we used the MTX from the docked *L. casei* complex.

A comparison of the site sequences for the three forms of DHFR analyzed in this study is given in Table 4. For the substrate binding site, only 15 of the 52 possible residues (29%) are equivalent, while in the NADPH binding site, only 14 out of the 59 postions (24%) are sequence-homologous. Thus, for the entire catalytic site region, there is only 24% sequence homology, i.e., 20 of the possible 85 residues.

Using the *chemical homology* scale, the agreement between the binding pockets of these three forms of DHFR is greatly increased to: (a) 58% for the substrate site (15 exact plus 15 backbone out of 52 total positions); (b) 56% for the NADPH binding site (14 exact plus 19 backbone out of total of 59 positions); and (c) 54% for the overall catalytic pocket (20 exact plus 26 backbone out of 85 total site residues).

The chemical homology of *Chicken* DHFR with either bacterial DHFR is ~60% for the catalytic site region, which is ~10% less than the agreement found between the *E. coli* and *L. casei* bacterial forms. Thus, we should expect that the free-energy profile for the steady-state kinetic pathway of the *Chicken* DHFR would not be as closely aligned as the two bacterial systems. Experimental testing of this conclusion is under way.

The major distinction in reaction profiles for these forms of DHFR is that

Table 4. Comparison[a] of DHFR active site residues from *E. coli*, *L. casei*, and *Chicken*.

Residue[b]	4	5	6	7	8	9	10	...	13	14	15
E. coli	L	i	A	**A**	L	A	V	...	V	i	**G**
L. casei	F	l	W	**A**	Q	N	R	...	L	l	**G**
Chick	S	i	V	**A**	V	C	Q	...	G	i	**G**

	16	17	18	19	20	21	22	...	23	24	25
E. coli	M	E	N	A	m	**P**	**W**	...	n	**L**	p
L. casei	K	D	G	H	l	**P**	**W**	...	h	**L**	p
Chick	K	D	G	N	l	**P**	**W**	p	p	**L**	r

	26	27	28	29	30	31	32	33	34	35	36
E. coli	A	d	l	a	w	**F**	k	R	N	**T**	l
L. casei	D	d	l	h	y	**F**	r	A	Q	**T**	v
Chick	N	e	y	k	y	**F**	q	R	M	**T**	s

	37	...	40	...	42	43	44	45	46	47	48
E. coli	d	...	g	...	m	**G**	r	h	**T**	W	e
L. casei	g	...	m	...	v	**G**	r	r	**T**	Y	e
Chick		...		...	m	**G**	k	k	**T**	W	f

	49	50	51	...	52	53	54	55	56	57	...
E. coli	**S**	i	g	...	**R**	**P**	**L**	p	G	**R**	...
L. casei	**S**	f	p	k	**R**	**P**	**L**	p	E	**R**	...
Chick	**S**	i	p	n	**R**	**P**	**L**	k	D	**R**	...

	61	62	63	64	65	66	67	68	...	74	75
E. coli	i	**L**	s	s	q	p	G	t	...	w	v
L. casei	v	**L**	t	h	q	e	D	y	...	v	v
Chick	v	**L**	s	r	e	l	K	e	...	l	s

Table 4. (continued)

	76	77	78	79	...	92	93	94	95	96
E. coli	K	S	v	D	...	m	V	i	**G**	**G**
L. casei	H	D	v	A	...	v	I	a	**G**	**G**
Chick	K	S	l	D	...		I	v	**G**	**G**

	97	98	99	100	101	102	103	...	111	112
E. coli	G	r	v	y	E	q	f	...	Y	L
L. casei	A	q	i	f	T	a	f	...	L	V
Chick	T	a	v	y	K	a	a	...	F	V

	113	114	115	...	122	123	124	125	...	153
E. coli	**T**	**H**	i	...	**D**	**T**	**H**	f	...	f
L. casei	**T**	**R**	l	...	**D**	**T**	**K**	m	...	y
Chick	**T**	**R**	i	...	**D**	**T**	**F**	f	...	f

[a] Bold upper case indicates homologous, lower case indicates nonhomologous, and nonbold upper case indicates backbone chemical homology, as described in Section 3.1.3 of the text.

[b] Numbered for the *E. coli* form.

for *Chicken* DHFR, the rate of folate reduction to DHF is about 100 times the rate for the two bacterial species.[20] This difference in rate may be related to the subtle differences between the bacterial and avian site homologies, as will be expanded upon in Section 8.

4. Simulation Strategies

Because of the difficulty in obtaining detailed structural information about DHFR substrate cofactor complexes and the difficulty of obtaining these data for a sequence of mutated systems, we have used molecular mechanics and molecular dynamics simulations to predict structures appropriate for solvated complexes at room temperature.[23] This procedure involves describing the energy of the ~3500 atom system (including solvent) as a superposition of two-body, three-body and four-body interactions (the force field). Using this force field the forces are calculated for each geometry and Newton's equations are solved to predict the dynamics.

For these studies we used the AMBER force field,[24] and the molecular dynamics and energy minimization calculations were carried out with the BIOGRAF[20] molecular simulation program (which has the dynamics integrated with three-dimensional graphics displays). This marriage of calculational abilities with visual ease provides the tool necessary for theoretical investigations of biochemical systems.

Specifics of the approaches used to define the starting structures for the simulations are presented below. This discussion is followed by a brief presentation of the force field used in the calculations and a discussion of the strategies developed to preceed from initial enzyme complex to equilibrated form. The results and conclusions from each set of simulations are presented and discussed in detail in Sections 5 through 8.

4.1 Models for Starting Complexes

As will be discussed, such molecular simulations are quite useful in predicting changes in structure that are due to mutations or to binding of substrates or

other molecules. However, these techniques are not yet adequate for predicting the optimal folding of the complete protein. Consequently it is essential to use an initial starting structure, which defines the overall protein folding. A complete experimental crystal structure would provide an ideal starting structure for the calculations. Unfortunately, only partial experimental structural data are available, as outlined in Section 2. Consequently, a number of manipulations and assignments were necessary to generate the initial structure data sets required for the molecular simulations.

For the studies presented here, we constructed DHFR binary and ternary complexes, suitably solvated. In some studies, alternate conformations of bound substrate or inhibitor were required. Mutant protein structures were also needed. Some details of the procedures employed to add the missing pieces to the available experimental data are described in the following sections.

4.1.1 Proteins

The crystallographic coordinates for *E. coli* Dihydrofolate Reductase were obtained from Brookhaven Data Files as reported by Matthews *et al.*[5,6] For this form of the enzyme, the crystal structure file contained two independent protein molecules in the unit cell, designated as **A** and **B** (equivalent to 1 and 2, respectively, as reported in the literature). One methotrexate (MTX) inhibitor molecule is bound to each protein molecule in a one-to-one complex. The crystallographers indicate that chain **B** is preferred for structural comparisons because it is more complete and less perturbed by intermolecular contacts than is chain **A**. For these reasons, the simulations were done using chain **B** as the starting DHFR structure.

The completion of the DHFR starting structure involved 2 stages. The

Brookhaven files do not contain coordinates for the carboxylate groups of Glu-129 and Asp-131. The Replace facilities of BIOGRAF were used in conjunction with the reported atomic postions of residues 129 and 131 to generate initial coordinates for these carboxylate groups. The crystallographic file also included coordinates for alternate conformations (A and B) of His-45, Ser-64 and Asp-122. Matthews and coworkers indicate that in all cases the A form is believed to be the major conformer. In each case energy minimization calculations also indicated that the A form is lower in energy than the B form. Consequently, the A conformer was used for these studies.

4.1.2 Substrate and Inhibitor Orientation

The starting structures for either the bound substrate or the inhibitor simulations were based upon the crystallographic data for the MTX bound to DHFR. For the experimental form of bound MTX, referred to as MTX-xtl, the coordinates were used as reported. For the pteridine-flipped form of bound MTX, termed MTX-anti, the C6-C9 and C9-N10 bonds of MTX were rotated to produce a starting conformation with the pteridine ring positioned in the active site but with its face flipped 180° (Figure 9). To be consistent with experiment, both forms of bound MTX inhibitor were protonated at $N1^{25}$.

As with the MTX, two orientations of the ***bound substrate*** were considered. The DHF units were positioned based upon: (a) the crystal coordinates of the MTX and hence termed DHF-xtl, and (b) the modeled coordinates of the flipped pteridine ring form (DHF-anti). The chemical structure of the bound MTX was "mutated" into that of the natural substrate; i.e., the carbonyl oxygen at position 4 of the pteridine ring was converted to an amino substituent and the benzoylamino unit was methylated. To be consistent with experimental observations,

Figure 9. The two orientations for the bound pteridine ring studied using molecular simulations. The unflipped form corresponds to the conformation found in the crystal complexes (referred to as xtl in the text) and the flipped form is achieved by rotation about the C6-C9 and C9-N10 bonds (equivalent to the anti form).

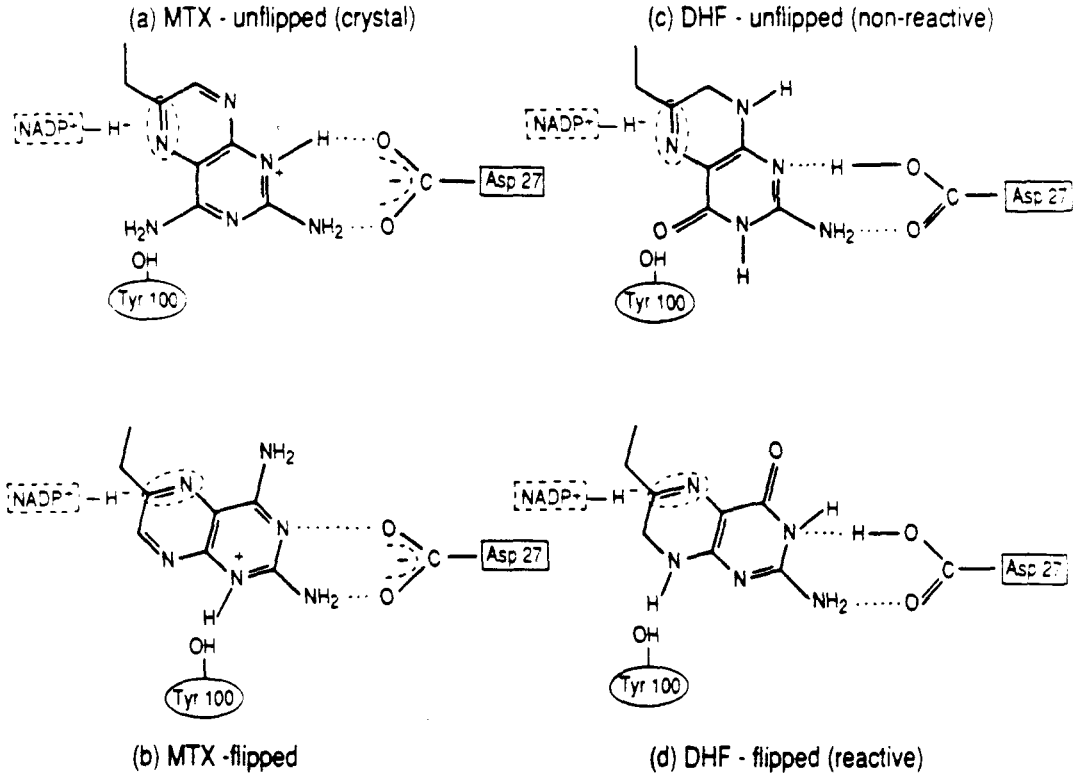

(a) MTX - unflipped (crystal)
(c) DHF - unflipped (non-reactive)
(b) MTX -flipped
(d) DHF - flipped (reactive)
NADP+
H-
Asp 27
Tyr 100
OH
NH2

Figure 9.

the carboxylate group of Asp-27 was protonated, while the pterin ring of the substrate was not protonated at the N1 position.

For studies of folate substrate with DHFR, the starting DHF-anti structure was altered to reflect a bound folate molecule. As with the dihydrofolate substrate, the carboxylate group of Asp-27 was protonated and the N1 position was unprotonated.[10,25]

4.1.3 Position of NADPH Cofactor

The coordinates for the *E. coli* bound NADPH were modeled based upon those from the *L. casei* ternary complex. The *L. casei* system docked with respect to the *E. coli* form (discussed in 3.1.1) provided starting coordinates for NADPH bound to the *E. coli* protein.

Addition of the NADPH cofactor to the *E. coli* binary complex necessitated two types of modifications to the full system. First, the solvent shell had to be corrected (*vide infra*). The water molecules lining the vacant NADPH site were deleted from the starting ternary form and appropriate counterions were positioned near the anionic sites of the NADPH.

The other changes involved the position of the protein loop holding the NADPH into its binding site. In the *E. coli* binary crystal structure, residues Glu-17 through Trp-22, are in an extended conformation open and away from the vacant cofactor binding site. The analogous amino acids in the *L. casei* ternary complex, however, are moved in toward the bound NADPH and assist in positioning the cofactor. Hence, the orientation of this fragment in the *E. coli* protein was modeled after that of the *L. casei*. This modification was accomplished by substituting for the coordinates of the *E. coli* residues the backbone atoms of the docked *L. casei* structure, thus creating a series of glycine amino acids. The side

chain positions were then determined by using the Replace facility of BIOGRAF in conjunction with molecular simulation (as discussed below) and the *L. casei* crystal structure.

4.1.4 Solvation

The Brookhaven data file for *E. coli* DHFR contains coordinates for the oxygen atoms of 197 water molecules associated with chain **B**. The coordinates for two ions were also reported, one Cl^- and one Ca^{+2}. The Cl^- was located between the side chains of His-45, Thr-46, Val-99 and Gln-102. It was maintained in the coordinate file. The Ca^{+2}, on the other hand, was located at the periphery of the protein structure. In our simulations we typically use Na^+ and Cl^- counterions, and since the Ca^{+2} was not sequestered by protein, it was removed from the starting structure file.

Additional water was added to the starting file with the Solvate option of BIOGRAF.[20] The diamond lattice type was used for the water addition. A grid spacing of 2.8 Å and a close contact cutoff distance of 2.8 Å were used. These values allowed for adequate H-bonding between solvent and protein as well as for the appropriate density for water. The MTX and eleven regions of the protein were defined and used for the segmented solvation. Solvent was added to a distance of 5 Å. This addition generated coordinates for an additional 458 waters, for a total of 655.

In such simulations it is important to consider electroneutral systems to avoid spurious energetics. This was achieved by positioning the appropriately charged counterions near cationic and anionic side chains. Certain water oxygens were selected to be changed into counterions near charged amino acid side chains, i.e., the side chains of Asp, Glu, Arg, and Lys. The positions were determined by

the coordinates of the next-nearest water. For the DHFR binary MTX simulation starting file, the results are listed in Table 5.

After the above manipulations, the final solvent for the binary complex simulation was composed of 619 waters (197 experimental and 422 modeled into the system), 23 Na^+ (all modeled waters converted to ions) and 13 Cl^- (1 experimental and 12 modeled waters converted to ions). For the ternary complexes, with DHF natural substrate, 21 waters were deleted from the NADPH binding site.

The starting structure for solvent within the active site region was examined using solvent accessible surfaces.[21,22] The solvent accessible molecular surfaces for the bound inhibitor or substrate, the NADPH cofactor, the active site of the protein, and the water molecules within 8 Å of the bound small molecules were generated (with a 1.4 Å probe radius). Visual examination of these molecular surfaces indicated that no vacancies or steric crowding were present. Thus an appropriate amount of solvent was present in the active site.

4.1.5 Site Mutations

The Replace portion of BIOGRAF was used to introduce site-specific mutants into the *E. coli* Dihydrofolate Reductase enzyme. This manipulation substitutes the desired mutant amino acid side chain for that of a preselected wild-type residue. The backbone atoms of the two amino acids, i.e., the new mutant and the native form, are superpositioned, while the side chain for the substituting residue is positioned to mimic the location of the natural amino acid. If the mutation involves loss of a substantial portion of the side chain, e.g., the *E. coli* Leu-54 to Gly-54 modification, appropriate solvent molecules are added to the site of the substitution.

Table 5. Those charged residues with neighboring HOH's changed into counter-ions.

Aspartate Residues

Number	HOH	Distance[a]
11	A-29	5.2236
27	MTX-N1	3.5537
37	E-13	6.0585
69	B-76	4.9038
70	B-78	5.5722
79	G-22	5.9090
87	E-25	4.6601
116	F-14	5.7514
122	A-76	5.1697
127	F-21	6.2717
131	F-17	4.5473
132	C-40	6.0440
142	F-1	5.1459
144	I-2	4.7541

Arginine Residues

Number	HOH	Distance[b]
12	C-52	4.9490
33	H-14	5.5488
44	B-74	4.5535
52	MTX-CD	5.5419
57	MTX-CT	4.1729
71	E-47	4.8016
98	A-81	5.3439
158	C-49	5.5145
159	X-2	4.8605

[a] To aspartate γ-carbon.

[b] To arginine guanadinium carbon.

Table 5. (continued)

Glutamate Residues

Number	HOH	Distance[a]
17	A-63	5.5756
48	J-6	5.9410
80	B-81	5.6403
90	Lys-38	
101	A-60	6.1143
118	A-16	5.7055
120	A-45	5.0320
129	C-46	6.6427
134	A-79	4.8605
139	C-15	5.6683
157	C-18	5.4142

Lysine Residues

Number	HOH	Distance[b]
32	E-9	4.8595
38	Glu-90	
58	E-39	4.9874
76	E-56	4.8419
106	G-25	4.9956
109	C-34	5.3602
154	C-7	5.5638

[a] To glutamate δ-carbon.

[b] To lysine terminal nitrogen.

4.1.6 Loop Insertions

The investigations of hybrid Dihydrofolate Reductase proteins depended upon the construction of "segment mutant" enzymes. This task was facilitated by using the BIOGRAF Protein Builder. The Insert feature permits segments of amino acids to be added to selected regions of a protein structure.

For the *E. coli* - *Chicken* hybrids discussed in Section 8, loop regions of the *E. coli* protein were extended, based upon the sequence and structure of the *Chicken* form. In loop-hybrid 1 (LH1), the *E. coli* segment Ile-Gly-Arg-Pro-Leu-Pro-Gly-Arg-Lys-Asn (sites 50 - 59) was replaced by the *Chicken* sequence Leu-Pro-Glu-Lys-Asn-Arg-Pro-Leu-Lys-Asp-Arg-Ile-Asn (positions 61 - 72). The coordinates of the atoms in this new loop were assigned from those of the docked *Chicken* structure file. Similarly, *E. coli* Leu-36 was changed to reflect the *Chicken* sequence Ser-Thr-Ser-His-Val-Glu-Gly-Lys-Gln (positions 39 through 47 in the wild-type *Chicken* structure) to yield loop-hybrid 2 (LH2).

4.2 Force Field

The molecular mechanics simulations were carried out with BIOGRAF, using the AMBER force field developed for describing proteins and nucleic acids[24]. This force field method involves harmonic potential functions for bond stretches and angle bends and cosine expansion potentials to describe torsions and inversions. The nonbond interactions are described with fixed charge electrostatic terms, with Lennard-Jones 12-6 potentials to account for van der Waals interactions, and a Lennard-Jones 12-10 potential to modulate the electrostatic terms involved in hydrogen bonding. A more detailed description of the force field potential functions is presented in Appendix I.

The water potential employed for these simulations is based upon the Jorgen-

son TIPS model[26] but uses the Pettitt modification for flexible waters.[27] This potential places charges on each atom ($Q_H = +\ 0.40$, $Q_O = -\ 0.80$) and uses Lennard-Jones 12-6 potentials on both O and H. See Appendix V for details of this water potential.

For the electrostatic energies, the partial charges for the amino acid residues were the values from AMBER. The atomic charges for methotrexate, dihydrofolate, folate, and NADPH were based upon values reported for analogous amino and nucleic acid fragments, and are given in Table 6.

4.3 Simulation Protocols

The series of calculations run on the solvated DHFR complexes consisted of two separate phases. The first phase was designed to bring the full system into equilibration. The second phase involved structural and energetic analyses. Each phase made use of molecular dynamics and energy minimization. A brief discussion of these techniques as well as their implementation in the equilibration and analysis of the DHFR complexes is presented below.

4.3.1 Energy Minimization and Molecular Dynamics

In energy minimization, the coordinates of the atoms in the molecular system are modified so as to decrease the energy of the entire system to a minimum. Given the set of 3N independent coordinates $\mathbf{x} = (x_1, x_2, x_3, \ldots, x)$ for the N atoms of the macromolecular system, we evaluate the potential energy function $V = V(\mathbf{x})$ and the force on each atom $F_i = -\nabla V_i = -\frac{\partial V}{\partial x_i}$. Conjugate gradient techniques are used to find the energy minimum rapidly. This method changes the 3N dimensional coordinate vector in a direction determined by the force but corrected by information from previous minimizations.

Such energy minimization is useful for refining molecular structures, i.e.,

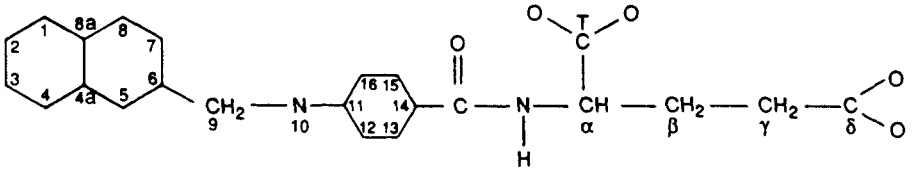
Table 6. Point charges for DHFR substrates and methotrexate inhibitor.

Atom[a]	MTX	DHF	THF
N1	-0.5000	-0.7020	-0.7020
HN1	0.4800	—	—
C2	0.8400	0.8420	0.8420
NA2	-0.6600	-0.7580	-0.7580
HNA2	0.3500	0.3300	0.3300
N3	-0.7000	-0.7460	-0.7460
HN3	—	0.3400	0.3400
C4	0.8100	0.7140	0.7140
NA4	-0.7900	—	—
OA4	—	-0.4590	-0.4590
HNA4	0.3400	—	—
C4A	-0.1500	-0.0880	-0.0880
N5	-0.7500	-0.7390	-0.5000
C6	0.7800	0.7800	0.1000
C7	0.1200	0.1200	0.1000
N8	-0.6500	-0.7500	-0.5000
HN8	0.3900	0.3400	0.3400
C8A	0.6000	0.4900	0.4900
C9	0.0800	0.0800	-0.0490
N10	-0.3800	-0.4440	-0.4440
CM	0.1000	—	—
HN10	—	0.3200	0.3200
C11	-0.1300	-0.1300	-0.1300
C12	0.0650	0.0650	0.0650
C13	0.0650	0.0650	0.0650
C14	-0.1300	-0.1300	-0.1300
C15	0.0650	0.0650	0.0650
C16	0.0650	0.0650	0.0650

Table 6. (continued)

Atom[a]	MTX	DHF	THF
C	0.5300	0.5300	0.5300
O	-0.5000	-0.5000	-0.5000
N	-0.5200	-0.5200	-0.5200
HN	0.2500	0.2500	0.2500
CA	0.2400	0.2400	0.2400
CT	0.5600	0.5600	0.5600
O1	-0.7800	-0.7800	-0.7800
O2	-0.7800	-0.7800	-0.7800
CB	0.0000	0.0000	0.0000
CG	-0.2000	-0.2000	-0.2000
CD	0.6200	0.6200	0.6200
OE1	-0.7100	-0.7100	-0.7100
OE2	-0.7100	-0.7100	-0.7100

[a] Labeled as in:



reducing steric crowding and adjusting bond lengths and angles to values near their optima. In addition, the relative energies of different conformations provide information about relative stabilities. However, for such complex molecules as enzymes, there can be several local minima. The conjugate gradient techniques lead to converged local minima of the potential surface closest to the starting point. Hence, the initial molecular structure greatly influences the final, minimized form. In such complex systems energy minimization calculations generally do not produce conformations that are very different from the starting, unrefined structure.

In molecular dynamics, the trajectory of the system at finite temperature (usually 300K in our studies) is calculated by solving Newton's equations for an interval of time. In such calculations, the 3N components of force on each of the N atoms are calculated for the geometry at time t_0 and Newton's equation is

$$\frac{\delta^2 \mathbf{x_i}}{\delta t^2} = -\frac{\nabla V(\mathbf{x_i})}{m_i} \tag{1}$$

solved, using the coordinates and velocities at time t_0, to predict the coordinates and velocities at time $t = t_0 + \delta$, where $\delta \sim 10^{-15}$ sec. Using the new coordinates at time $t_0 + \delta$, the 3N components of force are calculated and used to predict the coordinates at time $t_0 + 2\delta$, etc.

4.3.2 Initial Relaxations and Equilibration

The calculations used to bring the macromolecular systems into equilibrium proceeded in a stepwise manner that introduced additional segments of the complex with each phase of the process. In all cases, the initial structure was minimized to relieve the possibility of highly strained interactions. Once any repulsive van der Waals or high-energy covalent terms were reduced, equilibration was achieved by the use of molecular dynamics at T=300K.

The first portion of the structure to be relaxed was the solvent. Initially, the *added* water and counterions were allowed to equilibrate (made movable), while the structure of the protein, small molecules (substrate, inhibitor, cofactor) and experimental solvent was maintained in its initial form (kept fixed). The results from this equilibration were then used in the next step where *all* solvent molecules were movable, while the enzyme complex was held in a fixed postion.

Starting with this structure with relaxed solvent and nonrelaxed enzyme complex, we equilibrated the full system with the following series. First, the side chain units and small molecules bound to the catalytic pocket, along with all solvent molecules, were relaxed, using molecular dynamics, while the positions of all protein backbone atoms were held fixed in their original position. After this side chain equilibration, we then completely relaxed all atoms in the structure, using dynamics to form the starting structure for the molecular simulations.

4.3.3 Production Runs

We next evaluated the energetics and structures during a series of quenched-dynamics calculations, allowing all atoms (including solvent) to move freely and to equilibrate during such quenched-dynamics runs. The structures were periodically minimized to find the local structure corresponding to that point on the dynamics trajectory. For such minimizations, only the structure of the protein, substrate/inhibitor, and cofactor portions of the complex were permitted to adjust, while the solvent components were held in their equilibrated positions. These minimized structures were used for full and component energy analysis as well as for structural comparisons, as discussed in detail in the various sections below. After each such minimization, the dynamics were continued with the structure prior to the minimization.

5. Methotrexate Simulations

Molecular simulations on DHFR·MTX complexes were used to bench-mark the computational method and strategies. These calculations focused on three basic issues. First, is the force field adequate to produce the structure of DHFR complexes in good agreement with experiment? Second, do the simulations indicate the appropriate preference for the orientation of MTX binding? And third, is it possible to examine a smaller subset of the macromolecular assembly and retain the accuracy and reliability of the simulation results? The results of computational studies and analyses to address these questions are given below.

5.1 Full Structural Studies

The DHFR·MTX binary structure (constructed as described as in Section 4) was the starting point for calculations to test the validity of this force field and simulation methodology. The protein complex contained the MTX inhibitor bound in its crystallographic conformation. The macromolecule was equilibrated, as discussed above, and quench-dynamics simulations were carried out with energy minimization, at 0.5 ps intervals.

Root mean-square (RMS) comparison of the protein and MTX inhibitor structures during dynamics simulations and after structural refinement were compared with those of the crystallographic complex. The results of this analysis are given in the first column of Tables 7 and 8. Using heavy atoms only for the protein and inhibitor complex, the agreement between calculated structures and the crystal are within 0.90 Å. When only the protein backbone heavy atoms are considered, as illustrated in Figure 10, the RMS difference between theory and experiment is slightly improved to within 0.75 Å, indicating better agreement for the backbone of the structure than for the side chains. This is expected in light

Table 7. RMS deviations (heavy atoms only) of theoretical and experimental structures for full MTX/DHFR binary complexes.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
0.5Dyn	0.86	0.00							
1.0Dyn	0.87	0.41	0.00						
1.5Dyn	0.88	0.47	0.39	0.00					
2.0Dyn	0.90	0.52	0.44	0.40	0.00				
0.5Min	0.83	0.21				0.00			
1.0Min	0.85		0.20			0.28	0.00		
1.5Min	0.87			0.20		0.35	0.27	0.00	
2.0Min	0.89				0.20	0.42	0.34	0.27	0.00

Table 8. RMS deviations (backbone atoms only) of theoretical and experimental structures for full MTX/DHFR binary complexes.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
0.5Dyn	0.66	0.00							
1.0Dyn	0.69	0.30	0.00						
1.5Dyn	0.69	0.35	0.27	0.00					
2.0Dyn	0.72	0.39	0.33	0.29	0.00				
0.5Min	0.66	0.14				0.00			
1.0Min	0.67		0.13			0.23	0.00		
1.5Min	0.69			0.14		0.29	0.22	0.00	
2.0Min	0.72				0.14	0.34	0.28	0.22	0.00

Figure 10. The experimental (yellow protein backbone and red MTX) and theoretical (light blue) structures for the DHFR·MTX binary complex.

Figure 11. Active site comparison of the experimental (yellow protein and red MTX) and theoretical (light blue) structures for the DHFR·MTX binary complex.

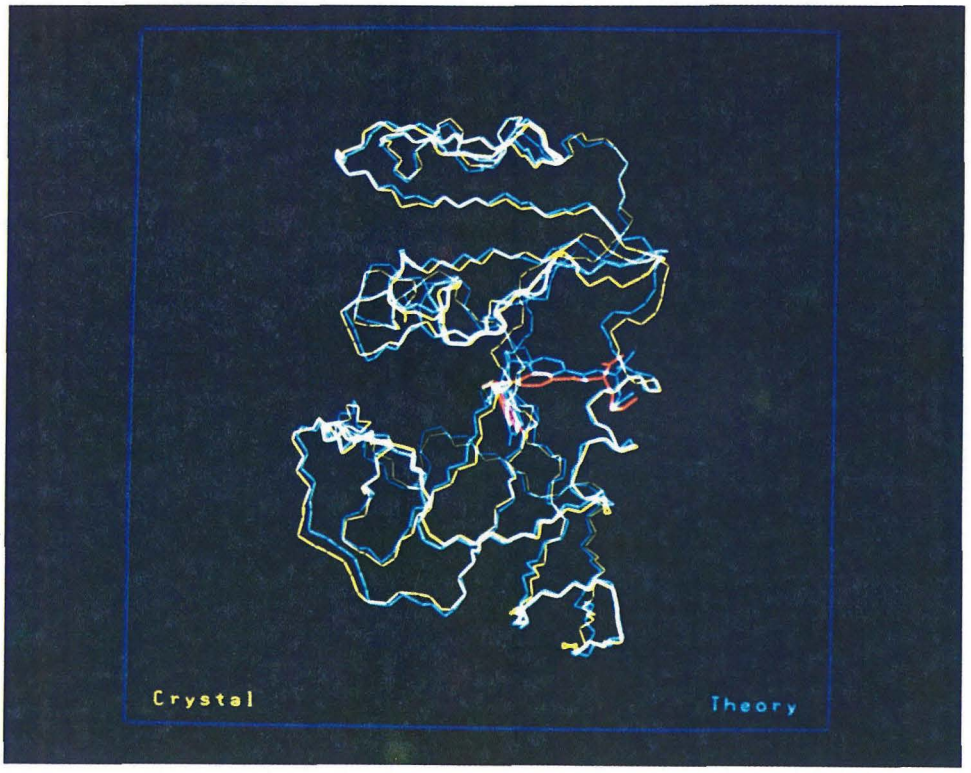


Figure 10.

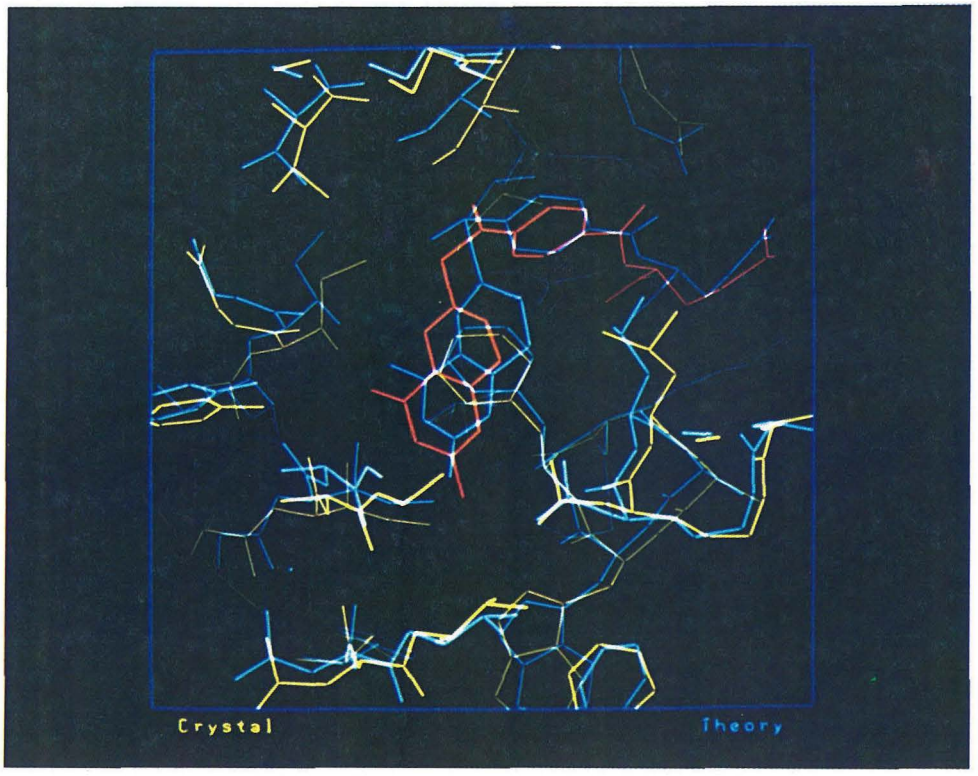


Figure 11.

of the greater rotational flexibilty associated with protein side chains relative to that of the backbone atoms. Figure 11 shows a comparison of the theoretical and experimental (crystal structure) active site regions including the amino acid side chains and MTX inhibitor.

RMS comparisons may also be used to provide a general indication of the changes in structure as the simulation proceeds. This is accomplished by determining the RMS difference between: (a) subsequent time segments of the molecular dynamics run, (b) appropriate pairs of dynamics and quenched structures, and (c) refined structures from neighboring intervals of the dynamics relaxation. The results from these studies are also given in Tables 7 and 8.

The RMS dynamics comparisons for these DHFR·MTX binary simulations indicate that the structures vary by ~0.40 Å from one 0.5 ps interval to the next (only ~0.29 Å for the backbone positions). This difference drops to ~0.27 Å, once the structures have been refined (~0.22 Å for the main-chain atoms). The differences between the dynamics forms and the corresponding minimized structures are quite small (~0.20 Å for the full protein and only ~0.14 Å for the residue backbones), indicating only subtle changes in structure during the energy minimization refinement process.

These results indicate reasonable agreement between theory and experiment (the enzyme structure has been refined to 1.7Å) and supports the utilization of the AMBER force field in simulations of Dihydrofolate Reductase complexes.

5.2 Orientation of Pteridine Ring in Bound MTX

As described in Section 2, there are two reasonable orientations for the pteridine ring of bound methotrexate. To study the relative structural and energetics for these two orientations, simulations were run with MTX in its flipped pterin

ring form, MTX-anti, to compare with the above MTX-xtl studies. The starting flipped-MTX protein complex was modeled as described in Section 4.1.2. Equilibration and production simulations, were identical to those run on the crystallographic form of bound MTX with DHFR. The results of various energetic analyses of these two MTX DHFR complexes are reported in Tables 9 to 13.

First, we present an analysis of the methotrexate inhibitor. Disregarding any contribution that is due to the enzyme or solvent, the simulations indicate that the energy of the crystallographic form of the bound MTX inhibitor is only slightly favored (by ~0.5 kcal) over that of the flipped form (see Table 9). The contribution to the energy from the torsional terms, E_ϕ, is ~50% lower in the crystal form than in the flipped form, while the angle and nonbond terms, E_θ, E_{vdW}, and E_{el} are more favorable for the flipped MTX. In all, the two conformations are equivalent, within the limits of the force field.

With the inclusion of the DHFR protein, the MTX conformational energy gap broadens. Evaluating the energy with the MTX as movable and the protein as fixed (thus including only nonbonded interactions between the enzyme and inhibitor) indicates that the crystallographic MTX is more stable than its flipped counterpart by ~10 kcal (Table 10). As in the case of the isolated inhibitors, the lower torsional energy of the xtl form dominates this preference, but in the presence of the protein, the nonbond energy terms favor the flipped form by only ~2 kcal. Thus, the enzyme stabilizes the xtl form of MTX by ~10 kcal more than it stabilizes the flipped form of this inhibitor.

In order to estimate the effect that the inhibitor has on the energy of the protein, we evaluate the energy of the complex with both the inhibitor and the protein macromolecule as movable. Under these circumstances, the energy of the

Table 9. Energetics for full MTX binary simulations: MTX movable, protein ignored.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	33.43	9.11	17.49	10.80	2.01	10.37	-16.36	613
1.0	45.20	10.16	21.10	16.05	3.16	12.34	-17.60	616
1.5	39.71	9.66	18.80	13.11	3.02	11.29	-16.17	617
2.0	42.96	10.20	19.50	14.33	2.88	11.17	-15.12	617
AVG	40.33	9.78	19.22	13.57	2.78	11.29	-16.31	616

MTX-anti.

0.5	39.80	11.41	14.84	25.89	3.77	4.77	-20.88	622
1.0	40.64	12.08	13.70	28.92	3.67	4.08	-21.82	619
1.5	41.32	11.23	14.77	26.78	3.89	4.42	-19.78	614
2.0	41.53	12.00	15.32	26.38	3.65	4.60	-20.42	613
AVG	40.82	11.68	14.66	26.99	3.75	4.47	-20.73	617

Table 10. Energetics for full MTX binary simulations: MTX movable, protein fixed.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	-159.98	9.11	17.49	10.80	2.01	-30.28	-169.12	9669
1.0	-154.95	10.16	21.10	16.05	3.16	-28.52	-176.90	9654
1.5	-156.54	9.66	18.80	13.11	3.02	-27.46	-173.67	9583
2.0	-155.50	10.20	19.50	14.33	2.88	-27.28	-175.13	9599
AVG	-156.74	9.78	19.22	13.57	2.78	-28.39	-173.71	9626

MTX-anti.

0.5	-154.18	11.41	14.84	25.89	3.77	-33.90	-176.19	9529
1.0	-139.62	12.08	13.70	28.92	3.67	-34.13	-163.86	9601
1.5	-142.85	11.23	14.77	26.78	3.89	-34.28	-165.25	9680
2.0	-151.42	12.00	15.32	26.38	3.65	-35.29	-173.49	9719
AVG	-147.02	11.68	14.66	26.99	3.75	-34.40	-169.70	9632

crystallographic form is favored by ~145 kcal, with dominant differences arising from the torsional, inversion, and electrostatic components (see Table 11). This energy difference appears rather large; however, this difference is less than 1 kcal per amino acid residue of the enzyme.

Another way of examining the effect of the MTX on the protein matrix is to consider only the site region of the protein, i.e., those residues in the enzyme with any atom within 8 Å of the bound MTX, and its interaction with the inhibitor. The energetic results of evaluating such a system are presented in Table 12. For the treatment with both the inhibitor and site portion of the protein movable, the MTX-xtl complex is once again energetically favored over the anti form by ~100 kcal. As above, the torsional and inversion terms contribute significantly to this difference, and the nonbond interations continue to favor the crystallographic form.

To extract from such energy differences the amount that is due solely to the energy of the site region of the protein, it is necessary to evaluate the interaction energy of inhibitor with the site, i.e., the energy of the movable MTX in the presence of the fixed site. Such an analysis indicates that the crystallographic conformation is ~10 kcal better stabilized by the enzyme site than is the flipped form. Thus, the energy of the XTL-site region is ~90 kcal more stable than the energy of the flipped-site region.

The role of solvent in the energetic differences of these complexes was also examined in Table 13. This table contains the energetic information for the cases with full DHFR·MTX binary complex (as movable atoms) immersed in its aqueous surroundings (fixed atoms). The inclusion of the fixed solvent contributes to the nonbond energy through van der Waals and electrostatic terms. The overall energy difference between the two forms of the inhibitor-bound complex is

Table 11. Energetics for full MTX binary simulations: MTX and protein movable.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	-1808.70	189.56	485.49	468.38	70.20	-768.88	-2253.45	165899
1.0	-1813.22	190.50	492.13	467.09	70.61	-768.56	-2265.00	166400
1.5	-1811.18	195.86	497.38	465.38	74.68	-757.69	-2287.10	166740
2.0	-1830.38	194.93	486.62	469.62	74.89	-747.56	-2308.37	167256
AVG	-1815.87	192.71	490.36	467.62	72.60	-760.67	-2278.48	166574

MTX-anti.

0.5	-1713.30	191.22	463.19	507.36	130.73	-773.20	-2232.60	165224
1.0	-1660.72	190.35	490.49	502.82	133.12	-756.72	-2220.77	166469
1.5	-1652.94	195.84	489.04	503.32	134.36	-744.16	-2231.34	167196
2.0	-1656.55	206.16	477.36	501.68	136.64	-746.34	-2232.05	167492
AVG	-1670.88	195.89	480.02	503.80	133.71	-755.11	-2229.19	166595

Table 12. Energetics for full MTX binary simulations: MTX and site region movable, remainder of protein ignored.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	-600.15	97.24	206.53	212.97	31.71	-302.40	-846.21	65257
1.0	-597.37	100.08	209.89	209.77	30.85	-301.52	-846.43	65217
1.5	-597.22	104.12	212.52	196.37	31.48	-297.88	-843.84	65187
2.0	-599.19	100.78	210.88	206.72	33.95	-292.72	-858.79	65378
AVG	-598.48	100.56	209.96	206.46	32.00	-298.63	-848.82	65260

MTX-anti.

0.5	-521.21	98.12	199.99	251.65	62.57	-287.49	-846.06	64610
1.0	-495.19	97.39	208.98	252.22	61.20	-285.23	-829.75	65119
1.5	-491.03	97.80	200.53	259.54	67.00	-279.66	-836.25	65136
2.0	-487.82	105.05	202.05	254.45	64.25	-273.30	-841.00	65349
AVG	-498.81	99.59	203.06	254.47	63.76	-281.42	-838.27	65054

Table 13. Energetics for full MTX binary simulations: enzyme complex movable, solvent fixed.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	-22907.58	189.56	485.49	468.38	70.20	-993.11	-23128.09	363957
1.0	-23208.60	190.50	492.13	467.09	70.61	-978.29	-23450.64	365128
1.5	-23270.86	195.86	497.69	465.38	74.68	-971.03	-23533.44	366201
2.0	-23378.22	194.93	486.11	469.62	74.89	-963.18	-23640.59	366690
AVG	-23191.32	192.71	490.36	467.62	72.60	-976.40	-23438.19	365494

MTX-anti.

0.5	-22990.99	191.22	463.19	507.36	130.73	-947.21	-23336.29	364501
1.0	-23159.19	190.35	490.49	502.82	133.12	-926.34	-23549.63	366734
1.5	-23295.09	195.84	489.04	503.32	134.36	-912.65	-23704.99	367858
2.0	-23403.78	206.16	477.36	501.68	136.64	-909.73	-23815.90	368148
AVG	-23212.26	195.89	480.02	503.80	133.71	-923.98	-23601.70	366810

~20 kcal, favoring the noncrystallographic, flipped form. This difference may be due to the different surface energies for our final shell of solvent (which could be eliminated by using periodic boundary conditions). Hence, while the solvent will be included in all simulations, the energetic analysis will focus only on the nonsolvent portion of the complexes.

The substrate binding pocket of DHFR is able to accommodate either the xtl or the anti conformation of the pteridine ring, as illustrated in Figures 12 through 14. After simulated relaxation and equilibration, a comparison of the binding of the MTX-xtl and MTX-anti forms, shown in Figure 14, indicates that: (a) the pteridine rings, initially coplanar, have adopted orientations that result in a relative twist of $\sim 45°$; (b) the phenyl rings, which occupied identical positions at the start of the simulations, have rotated and now exhibit a twist of $\sim 30°$; and (c) the glutamate fragments maintained their structural agreement. The amino acid residues in the active site adjust their positions to accommodate the two forms of inhibitor complexation. The conformation of the Phe-31 side chain in the MTX-anti structure is rotated $\sim 40°$, relative to its location in the MTX-xtl system, requiring the $C\delta_1$ center of Ile-94 to be displaced by ~1.5Å. The structural reorganization associated with the other active site side chains is typically 1Å, or less. The carboxylate group of Asp-27 is capable of salt-bridging to the MTX-xtl pteridine ring at the N1 and NA2 positions, with oxygen-to-nitrogen separations of ~3.0Å, as illustrated in Figure 12. This electrostatically favorable interaction is disrupted in the MTX-anti binding conformation (see Figure 13).

The energetic examination of the DHFR inhibitor species in the manner described above illustrates another interesting possibility. The energetic data for the movable MTX in a fixed protein "solvent" are remarkably similar for the two

Figure 12. The equilibrated DHFR active site region with MTX (red) bound in the xtl form. Phe-31 is visible to the right of the inhibitor's phenyl ring, Leu-54 appears above the pteridine ring, and Asp-27 is shown salt-bridged to the pteridine ring at the far right of the photo.

Figure 13. The equilibrated DHFR active site region with MTX (green) bound in the anti form. Phe-31 is visible behind the inhibitor's phenyl ring, Leu-54 appears above the pteridine ring, and the salt-bridge involving Asp-27 is no longer present.

Figure 12.

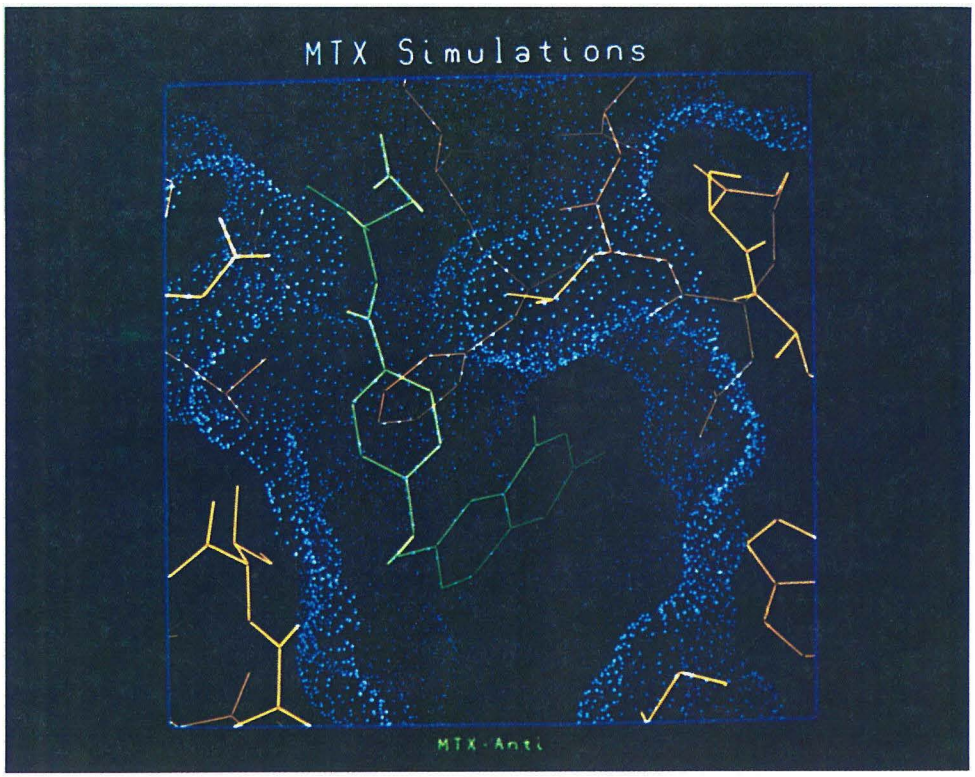


Figure 13.

Figure 14. The MTX xtl and anti equilibrated active sites depicting the different conformations adopted by the pteridine and benzyl groups.

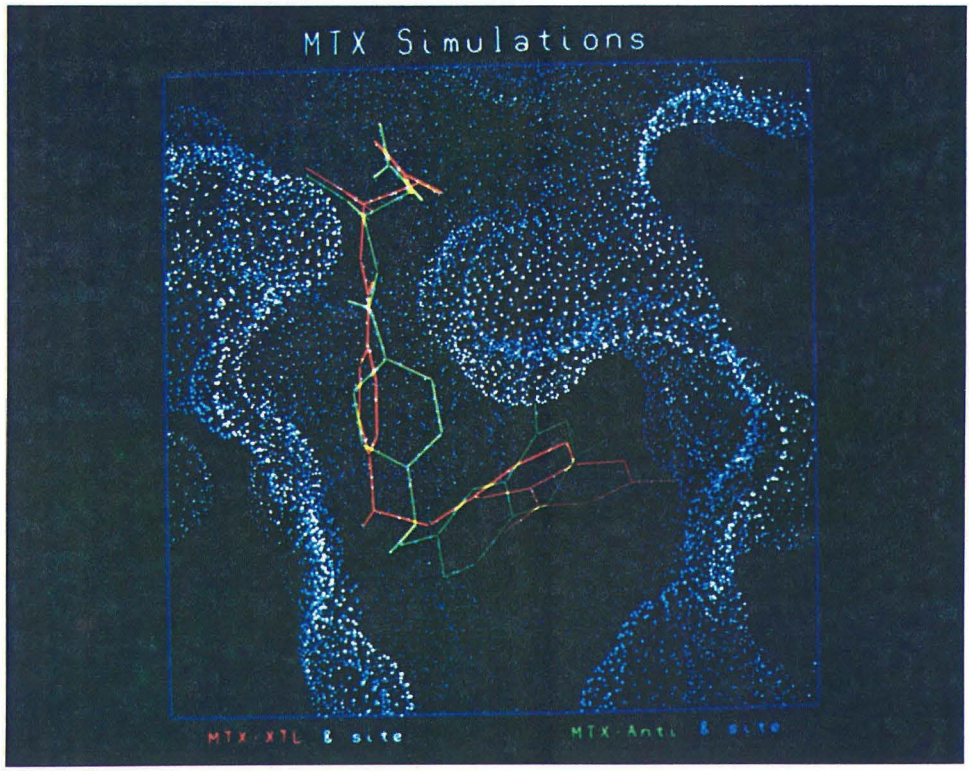


Figure 14.

situations considered, i.e., the full protein structure and just the site residues (see Tables 11 and 12). This agreement in the two energy profiles offers support for the concept of simulating these biochemical complexes with the active site movable and the remainder of the protein fixed. This approach to the calculations is advantageous in that it decreases the computational time by more than a factor of two. A series of computations were designed to further investigate the possibility of using site-region simulations for subsequent studies. The results are presented in the next section.

5.3 Site Region Studies

The substrate active site region of the *E. coli* DHFR was identified, based upon the position of the bound MTX inhibitor in the crystal structure. Work by Labanowski and coworkers[28] indicated that residues farther than 8 Å from the bound substrate did not contribute significantly to the nonbond energy terms. Hence, the substrate active site was defined to be those residues containing any atom within 8 Å of the bound MTX in either the crystal or the flipped forms. This definition resulted in an active site containing 72 amino acid residues, as listed in Table 14. The solvent molecules within 8 Å of the bound MTX and site protein were identified as site-solvent.

Site simulations were run for both the crystallographic and the flipped forms of bound MTX. For the molecular simulation energy expressions, bound inhibitor, site protein, and site solvent were defined to be movable, while the remainder of the enzyme and surrounding solvent were held fixed. The structure equilibration steps were identical to those for the full protein complex. The results from these site simulations were then compared with the corresponding runs where the entire enzyme complex was relaxed. The analysis consisted of structural integrity as

Table 14. Those DHFR amino acid residues within 8 Å of bound MTX.

Ser-3, Leu-4, Ile-5, Ala-6, Ala-7, Leu-8, Ala-9, Val-10
Ile-14, Gly-15
Asn-18, Ala-19, Met-20, Pro-21, Trp-22, Asn-23, Leu-24, Pro-25, Ala-26, Asp-27, Leu-28, Ala-29, Trp-30, Phe-31, Lys-32, Arg-33, Asn-34, Thr-35, Leu-36, Asp-37
Val-40, Ile-41, Met-42, Gly-43, Arg-44, His-45, Thr-46, Trp-47, Glu-48, Ser-49, Ile-50, Gly-51, Arg-52, Pro-53, Leu-54, Pro-55, Gly-56, Arg-57, Lys-58, Asn-59
Met-92, Val-93, Ile-94, Gly-95, Gly-96, Gly-97, Arg-98
Tyr-100, Glu-101
Leu-110, Tyr-111, Leu-112, Thr-113, His-114, Ile-115, Asp-116
Phe-125
Tyr-151, Cys-152, Phe-153, Lys-154, Ile-155

well as energetic comparisons.

The results from the structural analysis of the site simulations and comparisons with those structures form the full enzyme calculations are given in Tables 15 through 24. The RMS studies of the site simulations heavy and main chain atoms (comparing only those in the active site region), given in Tables 15 and 16, are quite similar to the results obtained for both the entire protein complex (given in Tables 7 and 8) and the site region only (see Tables 17 and 18) of the fully relaxed systems. Thus, the active site, as a whole, is not grossly perturbed in structure when performing simulations that use boundary conditions of the nonsite portion of the enzyme.

The possibility remained, however, that the amino acids adjacent to the fixed regions of the structure, referred to as site termini, might be distorted. To determine if this was the case, RMS structural comparisons were made of just the termini residues. The results, itemized in Tables 19 through 22, indicate that the positions of the site terminal amino acids were comparable to those of the full site region and the full protein. The forces on the atoms in the termini residues were also examined (see Table 23). When the remainder of the protein was ignored for the force evaluation, the values calculated from the structures produced from the site simulations agreed quite well with those forces acting on the same atoms in the files from the full protein relaxation. Similar agreement between site and full simulations was found when the nontermini amino acids of the protein were held fixed for the force analysis.

Direct comparisons of the corresponding active site structures from the site and full simulations are given in Table 24. These RMS values indicate that: (a) there is good overall agreement in the structure of the active site region when just the active site is allowed to relax, and (b) the terminal residues are not distorted

Table 15. RMS deviations (heavy atoms only) of theoretical and experimental site structures for site MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.76	0.00							
1.0Dyn	0.86	0.49	0.00						
1.5Dyn	0.92	0.57	0.45	0.00					
2.0Dyn	0.92	0.58	0.51	0.37	0.00				
0.5Min	0.74	0.23				0.00			
1.0Min	0.84		0.19			0.40	0.00		
1.5Min	0.90			0.19		0.49	0.34	0.00	
2.0Min	0.92				0.20	0.54	0.40	0.24	0.00

Table 16. RMS deviations (backbone atoms only) of theoretical and experimental site structures for site MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.58	0.00							
1.0Dyn	0.69	0.35	0.00						
1.5Dyn	0.75	0.41	0.30	0.00					
2.0Dyn	0.76	0.44	0.33	0.25	0.00				
0.5Min	0.58	0.15				0.00			
1.0Min	0.68		0.14			0.31	0.00		
1.5Min	0.74			0.13		0.36	0.25	0.00	
2.0Min	0.75				0.13	0.41	0.29	0.16	0.00

Table 17. RMS deviations (heavy atoms only) of theoretical and experimental site structures for full MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.85	0.00							
1.0Dyn	0.88	0.43	0.00						
1.5Dyn	0.85	0.46	0.39	0.00					
2.0Dyn	0.88	0.49	0.39	0.38	0.00				
0.5Min	0.83	0.21				0.00			
1.0Min	0.85		0.21			0.28	0.00		
1.5Min	0.85			0.19		0.35	0.25	0.00	
2.0Min	0.87				0.19	0.38	0.28	0.23	0.00

Table 18. RMS deviations (backbone atoms only) of theoretical and experimental site structures for full MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.65	0.00							
1.0Dyn	0.66	0.30	0.00						
1.5Dyn	0.64	0.34	0.24	0.00					
2.0Dyn	0.67	0.39	0.29	0.27	0.00				
0.5Min	0.61	0.14				0.00			
1.0Min	0.64		0.13			0.22	0.00		
1.5Min	0.65			0.13		0.29	0.20	0.00	
2.0Min	0.67				0.14	0.33	0.24	0.22	0.00

Table 19. RMS deviations (termini heavy atoms only) of theoretical and experimental site MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.59	0.00							
1.0Dyn	0.71	0.41	0.00						
1.5Dyn	0.74	0.44	0.29	0.00					
2.0Dyn	0.72	0.43	0.34	0.26	0.00				
0.5Min	0.59	0.19				0.00			
1.0Min	0.71		0.18			0.32	0.00		
1.5Min	0.73			0.15		0.36	0.19	0.00	
2.0Min	0.72				0.16	0.39	0.23	0.15	0.00

Table 20. RMS deviations (termini backbone atoms only) of theoretical and experimental site MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.46	0.00							
1.0Dyn	0.57	0.29	0.00						
1.5Dyn	0.58	0.31	0.22	0.00					
2.0Dyn	0.58	0.32	0.26	0.16	0.00				
0.5Min	0.48	0.15				0.00			
1.0Min	0.57		0.16			0.24	0.00		
1.5Min	0.58			0.10		0.27	0.13	0.00	
2.0Min	0.58				0.11	0.29	0.17	0.10	0.00

Table 21. RMS deviations (termini heavy atoms only) of theoretical and experimental full MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.82	0.00							
1.0Dyn	0.81	0.42	0.00						
1.5Dyn	0.82	0.45	0.36	0.00					
2.0Dyn	0.80	0.51	0.42	0.44	0.00				
0.5Min	0.81	0.22				0.00			
1.0Min	0.80		0.19			0.26	0.00		
1.5Min	0.81			0.19		0.31	0.24	0.00	
2.0Min	0.81				0.21	0.38	0.31	0.28	0.00

Table 22. RMS deviations (termini backbone atoms only) of theoretical and experimental full MTX/DHFR simulations.

		Dynamics				Minimized			
	XTL	0.5	1.0	1.5	2.0	0.5	1.0	1.5	2.0
XTL	0.00								
0.5Dyn	0.63	0.00							
1.0Dyn	0.62	0.30	0.00						
1.5Dyn	0.58	0.32	0.23	0.00					
2.0Dyn	0.63	0.38	0.32	0.33	0.00				
0.5Min	0.62	0.15				0.00			
1.0Min	0.61		0.12			0.22	0.00		
1.5Min	0.59			0.13		0.23	0.18	0.00	
2.0Min	0.64				0.13	0.32	0.29	0.29	0.00

Table 23. Force acting on atoms of termini residues.

Simulation	Termini of Active Site			
	Site Series with Remainder of Protein:		Full Series with Remainder of Protein:	
	Ignored	Fixed	Ignored	Fixed
0.5Dyn	26.96	29.62	27.86	29.96
1.0Dyn	29.26	32.63	26.69	28.67
1.5Dyn	25.91	28.19	22.81	24.82
2.0Dyn	27.95	30.29	27.21	29.31
0.5Min	12.35	10.04	11.60	9.97
1.0Min	12.49	10.28	11.93	9.80
1.5Min	12.40	10.37	11.65	9.47
2.0Min	11.84	10.04	11.84	9.51

Table 24. RMS deviations from direct comparison of theoretical structures for full abd site MTX/DHFR binary complexes.

Simulation	Site Residues		Termini Residues	
	Heavy	Main	Heavy	Main
0.5Dyn	0.98	0.73	0.94	0.68
1.0Dyn	1.06	0.78	0.95	0.69
1.5Dyn	1.06	0.84	0.99	0.67
2.0Dyn	1.11	0.89	0.97	0.71
0.5Min	0.94	0.70	0.89	0.65
1.0Min	1.02	0.77	0.96	0.67
1.5Min	1.05	0.84	0.96	0.67
2.0Min	1.10	0.88	0.97	0.71

by the presence of the fixed boundary. Thus, limiting the scope of the simulation to just the active site region does not appear to introduce unreasonable structural features at the protein boundary regions.

The results of the energetic analysis on simulations of the DHFR·MTX site are given in Tables 25 through 27. The energetic profile of the MTX inhibitor (Table 25) was slightly different from that found simulating the full complex. Overall, the energy of the small molecule was of the same order of magnitude in both cases. In the site studies, the crystallographic form of MTX was ~8 kcal higher in energy than the flipped form of the bound inhibitor. Comparing the site MTX to full MTX energies indicates that: (a) the anti forms are essentially equivalent, and (b) the most significant difference between the two xtl complexes arises from the torsional component to the energy. Thus, it appears that the decrease in flexibility of the protein structure in the site simulations most noticeably affects the torsional degrees of freedom associated with the bound inhibitor.

The introduction of the active site protein as "solvent" to the MTX indicates that the crystallographic complex is the more stable one (numerical data in Table 26). The site MTX complex in the crystal conformation is ~7.5 kcal lower in energy than the corresponding flipped structure. Thus, the protein stabilizes the experimentally observed form of the inhibitor ~15 kcal better than it stabilizes the anti form. Hence, the site simulations indicate that the experimentally observed form of the protein inhibitor complex is lower in energy than that form with the flipped pterin ring. This conclusion is in agreement with the results found from the full protein MTX calculational studies and supports using the site simulation strategy in subsequent studies.

As a brief side note, the energy of the two conformations is approximately equivalent (within 2 kcals) when the aqueous environment of the active site is

Table 25. Energetics for site MTX binary simulations: MTX movable, protein ignored.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	49.17	8.69	19.03	28.53	3.17	9.37	-19.62	645
1.0	45.99	7.75	16.44	28.53	2.95	10.76	-20.45	641
1.5	49.09	8.16	21.96	24.75	3.67	10.54	-20.05	636
2.0	49.10	8.16	20.26	27.45	3.51	11.16	-21.83	637
AVG	48.34	8.19	19.42	27.32	3.33	10.46	-20.49	640

MTX-anti.

0.5	38.45	7.28	17.60	22.59	3.77	12.23	-25.02	608
1.0	40.58	6.21	15.20	25.70	3.79	13.60	-23.93	613
1.5	39.43	6.85	14.62	27.76	3.95	11.10	-24.85	608
2.0	41.21	9.19	14.75	26.59	3.46	10.14	-22.93	613
AVG	39.92	7.38	15.54	25.66	3.74	11.77	-24.18	611

Table 26. Energetics for site MTX binary simulations: MTX and site region movable, remainder of protein ignored.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	-385.52	99.67	281.73	254.64	70.28	-278.68	-813.16	64275
1.0	-400.18	102.54	264.13	273.17	71.34	-276.84	-834.51	64058
1.5	-392.77	103.89	263.16	275.19	72.83	-278.42	-829.43	64734
2.0	-363.33	105.55	276.28	283.72	75.69	-266.88	-837.70	64747
AVG	-385.45	102.91	271.33	271.68	72.54	-275.21	-828.70	64454

MTX-anti.

0.5	-385.84	98.02	255.10	272.43	69.21	-261.48	-819.11	64534
1.0	-364.80	94.73	261.31	276.48	69.36	-260.95	-805.73	64270
1.5	-392.19	96.14	260.90	273.90	68.67	-273.58	-811.22	64414
2.0	-368.72	98.80	245.94	291.69	68.66	-259.85	-813.96	64099
AVG	-377.89	96.92	255.81	278.63	68.98	-263.97	-812.51	64329

added to the site complex (Table 27). Thus, as in the full structure simulations, the water solvent modulates the energetic differences between the xtl and the flipped-inhibitor protein complexes.

Table 27. Energetics for site MTX binary simulations: MTX and site protein movable, site solvent fixed.

MTX-xtl.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
0.5	-7831.65	99.67	281.73	254.64	70.28	-353.73	-8184.24	148527
1.0	-7887.31	102.54	264.13	273.17	71.34	-358.17	-8240.32	148464
1.5	-7944.78	103.89	263.16	275.19	72.83	-353.14	-8306.72	149359
2.0	-8004.96	105.55	276.28	283.72	75.69	-342.90	-8403.30	149359
AVG	-7917.18	102.91	271.33	271.68	72.54	-351.99	-8283.65	148927

MTX-anti.

0.5	-7806.45	98.02	255.10	272.43	69.21	-328.39	-8172.82	149147
1.0	-7920.65	94.73	261.31	276.48	69.36	-332.34	-8290.18	148869
1.5	-7964.63	96.14	260.89	273.90	68.69	-341.93	-8322.32	149288
2.0	-7969.75	98.80	245.94	291.69	68.66	-324.87	-8349.97	148822
AVG	-7915.37	96.92	255.81	278.63	68.98	-331.88	-8283.82	148782

6. Dihydrofolate Simulations

Simulations were undertaken on the natural substrate ternary complex of *E. coli* Dihydrofolate Reductase. Because of the highly reactive nature of the ternary species, detailed experimental structural information on this native system is unavailable. Thus, molecular simulation provides some insights on the DHFR system that are not yet experimentally attainable. This new information includes the role(s) of specific amino acids in: (a) substrate binding, (b) product release, and/or (c) catalytic functionality. Also, the conformational preference of bound substrate, as described above for the methotrexate inhibitor, may be investigated.

The purpose of these studies was twofold. First, we sought an atomistic understanding of the role DHFR plays in catalyzing the reduction of dihydrofolate to tetrahydrofolate. This picture includes the way the protein brings the two reacting moieties, i.e., DHF and NADPH, together for catalysis, the orientation of the substrate's pteridine ring, and the positions of key amino acids in the equilibrated active site. Second, these investigations served as controls for subsequent studies on mutant forms of the protein, the results of which are described more fully in Section 7.

The simulations required *complete* ternary *E. coli* DHFR·DHF·NADPH complexes. These starting structures were obtained by: (a) modifying the *E. coli* binary crystallographic information to include the NADPH cofactor; (b) altering the conformation of residues 17 to 22, at the NADPH binding site, to "hug" the cofactor (as is the case for the analogous amino acids in the *L. casei* ternary complex); (c) generating the bound DHF substrate from the coordinates of the MTX inhibitor; (d) protonating the aspartate residue at position 27; and (e) including additional water and counterions to the solvent. The details of these manipulations are discussed above in Section 4. Thus, the active site region of the protein

contained 101 amino acids (as listed in Table 28) and 408 water molecules.

Two issues relevant to the simulations were: (a) the orientation of the DHF pteridine ring and (b) the site for protonation of Asp-27. With regard to the substrate conformation, two different forms were studied. These are analogous to those of the MTX discussed in the previous section. For the DHF simulations, the form of DHF structurally equivalent to the crystallographic MTX is referred to as DHF-xtl. This orientation of bound dihydrofolate does not give rise to the reduced tetrahydrofolate product, as discussed in Section 2.2. The other form of DHF, termed DHF-anti, is that conformation with the pteridine ring flipped 180 degrees relative to the DHF-xtl structure. It is this form that isotope studies show to be involved in catalysis. The concern associated with the protonation of Asp-27 focuses on identifying which of the carboxylate oxygens would carry the hydrogen. The issue becomes more complicated when alternate forms of the pteridine ring are involved; e.g., what may be best for the DHF-xtl system may not adequately accommodate the DHF-anti form. Thus, four different protein ternary complexes were simulated. These structures had: (a) DHF in its xtl or anti form and (b) either Asp-27 OD1 or Asp-27 OD2 protonated.

The simulation protocols used for the equilibration and production runs were those described above. The final molecular dynamics calculations were carried out for 10 ps of simulation time. The complexes were refined for structural and energetic analysis at 2 ps intervals. The results from the analyses are presented in Tables 29 to 33.

The energetics of the two reactive small molecules, i.e., the DHF and NADPH, indicates that the structures derived from the anti form of DHF are $\sim$10 kcal more stable than those from the DHF-xtl studies (see Table 29). This difference is primarily due to the energy difference between the forms of DHF

Table 28. Those DHFR amino acid residues within 8 Å of bound DHF and NADPH.

Ser-3, Leu-4, Ile-5, Ala-6, Ala-7, Leu-8, Ala-9, Val-10
Arg-12, Val-13, Ile-14, Gly-15, Met-16, Glu-17, Asn-18, Ala-19, Met-20, Pro-21, Trp-22, Asn-23, Leu-24, Pro-25, Ala-26, Asp-27, Leu-28, Ala-29, Trp-30, Phe-31, Lys-32, Arg-33, Asn-34, Thr-35, Leu-36, Asp-37
Val-40, Ile-41, Met-42, Gly-43, Arg-44, His-45, Thr-46, Trp-47, Glu-48, Ser-49, Ile-50, Gly-51, Arg-52, Pro-53, Leu-54, Pro-55, Gly-56, Arg-57, Lys-58, Asn-59, Ile-60, Leu-62, Ser-63, Ser-64, Gln-65, Pro-66, Gly-67, Thr-68
Trp-74, Val-75, Lys-76, Ser-77, Val-78, Asp-79, Glu-80, Ala-81
Met-92, Val-93, Ile-94, Gly-95, Gly-96, Gly-97, Arg-98, Val-99, Tyr-100, Glu-101, Gln-102, Phe-103
Leu-110, Tyr-111, Leu-112, Thr-113, His-114, Ile-115, Asp-116
Val-119, Glu-120, Gly-121, Asp-122, Thr-123, His-124, Phe-125
Tyr-151, Cys-152, Phe-153, Lys-154, Ile-155

Table 29. Energetics for wild-type DHFR: DHF and NADPH movable, protein ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	111.33	28.79	84.47	60.26	10.67	6.98	-79.84	1853
4.0	121.87	34.64	77.16	68.52	14.03	3.52	-76.00	1833
6.0	121.88	28.66	78.88	72.06	19.22	5.06	-82.00	1855
8.0	131.87	29.47	80.20	76.77	20.87	5.26	-80.69	1853
10.0	117.47	29.98	76.38	70.27	18.47	5.01	-82.65	1843
AVG	120.88	30.31	79.42	69.58	16.65	5.17	-80.24	1847

DHF-anti Conformer; OD2 protonated.

2.0	115.15	31.05	85.44	73.13	14.36	8.72	-97.55	1863
4.0	130.84	36.94	93.87	73.12	12.82	8.30	-94.22	1857
6.0	133.29	35.12	95.67	71.22	14.82	10.42	-93.97	1857
8.0	118.01	37.34	83.15	72.32	13.63	7.42	-95.87	1851
10.0	120.16	36.85	87.89	69.81	12.69	7.14	-94.22	1861
AVG	123.49	35.46	89.20	71.92	13.66	8.40	-95.17	1858

DHF-xtl Conformer; OD1 protonated.

2.0	115.54	29.41	80.24	72.94	11.12	7.72	-85.87	1913
4.0	133.32	33.63	84.88	83.22	14.30	8.09	-90.79	1905
6.0	137.00	29.12	89.37	80.53	15.73	8.29	-86.03	1915
8.0	127.12	31.72	83.66	80.20	13.87	7.72	-90.03	1906
10.0	138.18	32.93	93.09	84.50	13.00	7.15	-92.48	1906
AVG	130.23	31.36	86.25	80.28	13.60	7.79	-89.04	1909

DHF-xtl Conformer; OD2 protonated.

2.0	126.68	36.43	97.51	62.83	12.08	9.89	-92.07	1835
4.0	143.54	38.92	98.28	71.85	13.35	10.88	-89.74	1840
6.0	129.45	36.35	91.60	69.99	10.47	12.07	-91.03	1845
8.0	129.60	32.46	96.13	66.38	11.46	13.57	-90.40	1847
10.0	136.75	37.00	94.84	68.87	11.76	11.94	-87.67	1837
AVG	133.20	36.23	95.67	67.98	11.82	11.67	-90.18	1841

(data given in Table 30), while the NADPH is essentially equivalent in the four cases (see Table 31). The stabilization gained from the two molecules interacting with each other, i.e., difference in energy between the two together and the sum of the individual energies, is greater for the OD1 protonation scheme (4.17 kcals anti form and 5.03 kcals xtl form) than for the OD2 structures (2.23 and 3.20 kcals, respectively).

With the addition of the active site protein, the energetic differences between the four cases split further (Table 32). The DHF-anti orientation with OD1 protonated is 38 kcals lower in energy than the OD2 protonated form. This difference in energy is attributable primarily to the angle components (three-body) of the energy. The DHF-anti OD1 system is also more stable than either of the DHF-xtl complexes by at least 15 kcals. Thus, energetically speaking, the ternary DHFR·DHF· NADPH complex favored is the DHF-anti OD1 form.

As found in the DHFR·MTX simulations, the DHFR active site will equilibrate to accommodate either the anti or xtl conformation of the substrate pteridine ring, as illustrated in Figure 15. After molecular simulation, a comparison of the binding of the DHF-anti and DHF-xtl forms indicates that: (a) the pteridine rings are essentially coplanar, unlike the results from the studies on the MTX binary complexes; (b) the phenyl rings have rotated only slightly and maintain a twist of $\sim 10°$; and (c) the glutamate fragments possess quite similar structures. The difference seen between the two conformations of bound substrate and bound inhibitor, i.e., relative coplanarity or twisting of the pteridine ring systems, are most likely due to the occupation state of the NADPH site. In the DHFR·MTX binary simulations, the cofactor site contains solvent molecules that permit some flexibility in the orientation of the inhibitor pteridine ring system. The cofactor cavities in the DHFR·DHF ternary studies, on the other hand, hold

Table 30. Energetics for wild-type DHFR: DHF movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	31.27	5.06	15.56	17.78	3.18	10.45	-20.75	538
4.0	46.07	5.75	15.46	24.47	7.24	9.66	-16.50	552
6.0	42.44	5.58	13.22	21.68	10.67	9.65	-18.36	550
8.0	56.62	5.86	16.94	28.70	12.13	9.75	-16.77	552
10.0	45.45	6.14	13.54	24.02	10.37	10.19	-18.80	545
AVG	44.37	5.68	14.94	23.33	8.72	9.94	-18.24	547

DHF-anti Conformer; OD2 protonated.

2.0	38.68	7.01	16.02	24.79	5.91	9.21	-24.25	542
4.0	37.72	6.04	16.43	25.22	5.78	8.07	-23.82	540
6.0	38.93	6.96	18.17	22.86	7.07	8.28	-24.41	547
8.0	38.33	7.95	17.02	25.44	5.80	7.20	-25.08	544
10.0	37.08	7.30	19.01	23.16	4.87	6.87	-24.12	547
AVG	38.15	7.05	17.33	24.29	5.89	7.93	-24.34	544

DHF-xtl Conformer; OD1 protonated.

2.0	47.67	5.84	14.83	25.81	4.61	7.75	-11.16	561
4.0	51.18	7.12	15.95	30.47	6.43	8.89	-17.67	562
6.0	46.56	5.78	15.49	26.13	5.01	9.01	-14.87	565
8.0	48.37	6.93	17.34	28.07	6.07	8.07	-18.11	561
10.0	53.02	7.68	21.17	29.93	5.05	7.33	-18.13	560
AVG	49.36	6.67	16.96	28.08	5.43	8.21	-15.99	562

DHF-xtl Conformer; OD2 protonated.

2.0	49.56	10.31	14.78	20.24	6.42	13.18	-15.36	551
4.0	55.50	7.85	12.57	29.47	6.12	13.60	-14.11	566
6.0	53.91	7.51	11.86	29.87	4.33	15.27	-14.94	566
8.0	51.91	6.99	12.86	26.37	4.90	15.58	-14.79	566
10.0	49.36	6.75	10.96	26.42	4.36	14.99	-14.11	564
AVG	52.05	7.88	12.61	26.47	5.23	14.52	-14.66	563

Table 31. Energetics for wild-type DHFR: NADPH movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	82.32	23.73	68.91	42.48	7.48	-2.52	-57.77	807
4.0	80.27	28.88	61.70	44.05	6.79	-3.65	-57.52	806
6.0	83.98	23.08	65.66	50.38	8.56	-2.25	-61.45	813
8.0	79.89	23.60	63.25	48.07	8.74	-2.04	-61.72	813
10.0	76.94	23.84	62.85	46.25	8.10	-2.48	-61.62	810
AVG	80.68	24.63	64.47	46.25	7.93	-2.59	-60.02	810

DHF-anti Conformer; OD2 protonated.

2.0	79.81	24.04	69.42	48.35	8.45	2.01	-74.46	829
4.0	95.67	30.90	77.45	47.90	7.04	2.70	-70.31	821
6.0	95.84	28.16	77.50	48.36	7.75	3.65	-69.58	820
8.0	81.42	29.39	66.13	46.88	7.84	2.10	-70.92	818
10.0	85.10	29.56	68.89	46.64	7.82	2.22	-70.03	821
AVG	87.57	28.41	71.88	47.63	7.78	2.54	-71.06	822

DHF-xtl Conformer; OD1 protonated.

2.0	72.42	23.57	65.41	47.12	6.51	2.50	-72.69	826
4.0	88.04	26.51	68.93	52.74	7.87	2.03	-70.05	824
6.0	94.33	23.34	73.88	54.40	10.72	1.96	-69.96	822
8.0	82.77	24.79	66.32	52.12	7.80	1.95	-70.21	826
10.0	91.96	25.25	71.92	54.57	7.95	2.57	-70.30	826
AVG	85.90	24.69	69.29	52.19	8.17	2.20	-70.64	825

DHF-xtl Conformer; OD2 protonated.

2.0	80.84	26.12	82.73	42.59	5.66	-1.21	-75.06	820
4.0	90.97	31.07	85.72	42.38	7.23	-1.42	-74.01	823
6.0	78.51	28.84	79.74	40.11	6.14	-1.90	-74.42	826
8.0	80.97	25.47	83.27	40.01	6.56	-0.67	-73.68	827
10.0	90.45	30.26	83.88	42.45	7.40	-1.54	-71.99	824
AVG	84.35	28.35	83.07	41.50	6.60	-1.35	-73.83	824

Table 32. Energetics for wild-type DHFR: DHF, NADPH, and site movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	-868.50	145.88	394.56	382.17	88.76	-456.41	-1423.46	104998
4.0	-854.83	153.92	371.11	390.96	97.66	-447.77	-1420.72	105483
6.0	-828.38	148.51	377.98	401.52	100.84	-437.72	-1419.49	105291
8.0	-793.08	150.98	381.51	422.27	101.00	-438.79	-1410.05	105282
10.0	-834.57	151.39	372.78	404.21	98.18	-438.13	-1422.98	105411
AVG	-835.87	150.14	379.59	400.23	97.29	-443.76	-1419.34	105293

DHF-anti Conformer; OD2 protonated.

2.0	-821.22	156.63	410.76	382.98	96.52	-447.06	-1421.04	104872
4.0	-796.61	162.99	411.83	397.34	96.60	-442.72	-1422.65	105031
6.0	-782.25	154.49	416.61	408.01	103.43	-431.06	-1433.73	105287
8.0	-791.61	154.84	396.80	414.54	100.82	-421.25	-1437.35	104854
10.0	-797.64	155.95	405.52	403.88	100.62	-433.24	-1430.38	105257
AVG	-797.87	156.98	408.30	401.35	99.60	-435.07	-1429.03	105060

DHF-xtl Conformer; OD1 protonated.

2.0	-845.92	155.91	402.59	381.61	95.01	-442.77	-1438.26	106945
4.0	-811.21	162.83	410.53	395.44	92.79	-430.44	-1442.35	106736
6.0	-833.40	147.88	402.68	395.28	93.93	-440.26	-1432.91	106728
8.0	-799.28	152.92	422.61	400.91	95.46	-441.60	-1429.57	107149
10.0	-813.74	149.75	421.25	399.53	92.78	-440.71	-1436.34	106704
AVG	-820.71	153.86	411.93	394.55	93.99	-439.16	-1435.89	106852

DHF-xtl Conformer; OD2 protonated.

2.0	-810.77	167.30	413.26	367.30	109.42	-426.27	-1441.77	105359
4.0	-786.88	164.73	397.55	399.98	103.78	-433.68	-1419.23	106131
6.0	-807.75	166.28	403.92	397.33	100.24	-447.39	-1428.12	106504
8.0	-840.14	162.33	397.79	376.35	94.74	-456.75	-1414.61	106842
10.0	-819.40	169.82	411.05	366.48	103.06	-452.48	-1417.33	107029
AVG	-812.99	166.09	404.71	381.49	102.25	-443.31	-1424.21	106373

Figure 15. A. The DHF-anti and DHF-xtl equilibrated active site regions, indicating the subtle differences that arise in the catalytic pocket because of the altered conformation of substrate binding.

Figure 15. B. The DHF binding site as the NADPH cofactor would see it. The nicotinamide ring of the NADPH, if shown, would be located just in front of the pteridine ring. This view clearly differentiates the two binding orientations of the substrate pteridine ring.

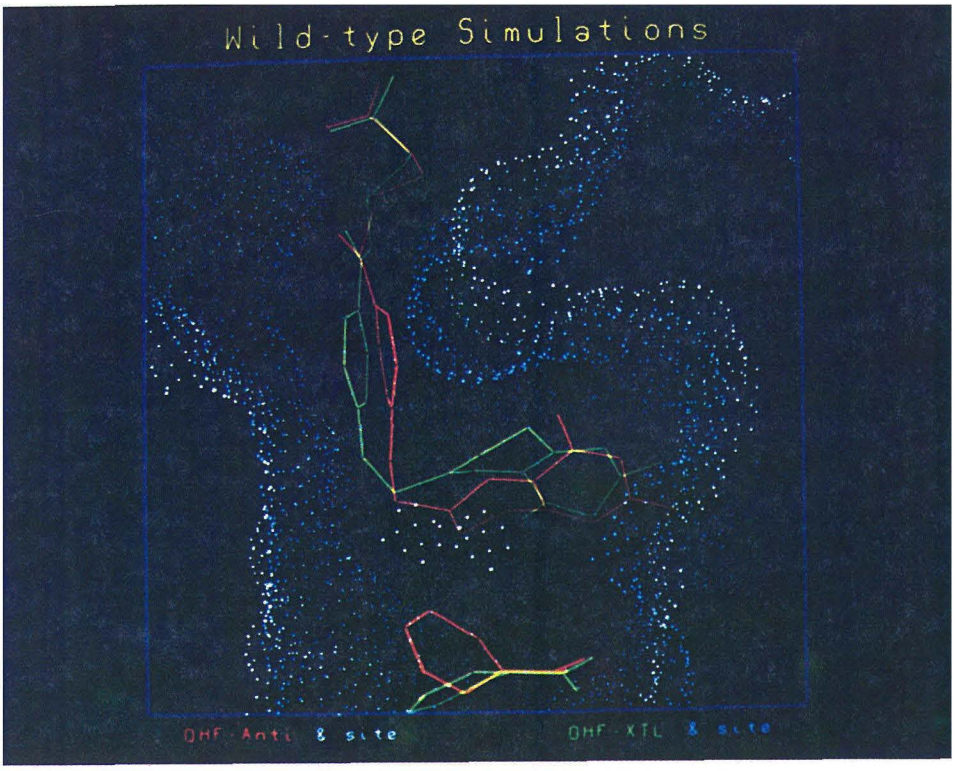


Figure 15. A.

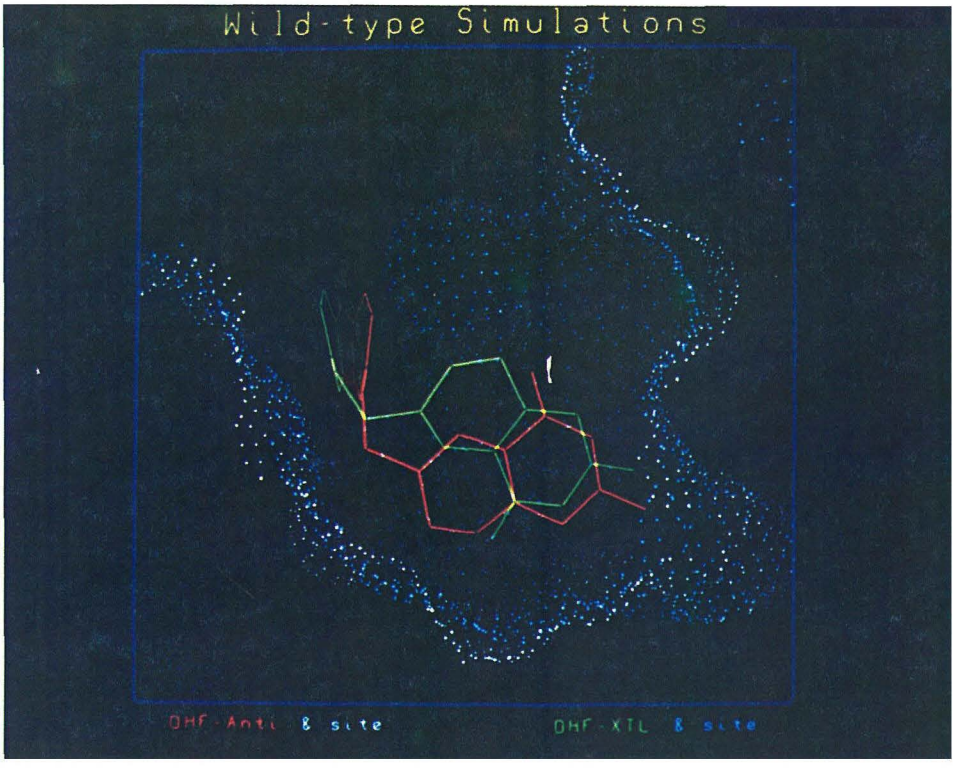


Figure 15. B.

the NADPH molecule and position its nicotinamide ring adjacent to the bound substrate, thus helping to maintain the conformation of the pteridine rings. The amino acid residues in the active site do adjust to accommodate the two binding modes of the substrate; however, the structural displacements are much smaller than those found for the MTX complexes.

A structural analysis of the two NADPH cofactor positions shows that: (a) the nicotinamide rings adopt conformations that place the plane of the rings approximately perpendicular to each other; (b) the positions of the nicotinamide ribose rings and the diphosphate fragments are quite similar; and (c) the adenosine units are slightly skewed. The altered conformations of the nicotinamide rings are most likely due to the change in binding orientation of the pteridine ring, as discussed above. The differences in conformation of the adenosine moieties appear at the solvent interface and hence do not disrupt the structural integrity of amino acids surrounding the bound cofactor.

One key structural feature for the DHFR·DHF·NADPH ternary complex is the distance associated with the hydride transfer. NC4 of the cofactor nicotinamide ring holds the hydride prior to catalysis, while C6 of the dihydrofolate's pteridine ring is the sight for deposition. The results of examining this hydride transfer distance in the simulated structures are presented in Table 33. It is indeed the case that this distance is shorter for the substrate that began the simulations in the anti form. With OD1 protonated, the wild-type hydride transfer distance is 3.63 Å, while the separation increases to 4.42 when OD2 carries the proton. Based upon these studies, the OD1 form is clearly favored for hydride transfer. For the xtl form of the substrate, the transfer distances are longer than those corresponding to the anti form (4.46 Å for OD1 and 5.55 Å for OD2). This longer distance is consistent with the observation that only the anti form is

Table 33. Wild-type DHFR: hydride transfer distances.

	DHF-anti		DHF-xtl	
	OD1	OD2	OD1	OD2
2.0	3.79	4.69	4.43	5.17
4.0	3.62	4.39	4.39	5.73
6.0	3.55	4.29	4.52	5.75
8.0	3.63	4.33	4.47	5.67
10.0	3.58	4.42	4.49	5.42
AVG	3.63	4.42	4.46	5.55

catalytic.

Both the energetic and distance analyses indicate that it is the OD1 protonated DHF-anti ternary complex that is the reactive form for wild-type DHFR. These studies indicate that a distance of 3.6Å separation of reaction centers in the equilibrated complex can lead to reactivity, whereas 4.4 Å does not.

7. Single-Site Mutations

Two amino acid sites in the active site of Dihydrofolate Reductase were targeted for investigation by both experimental and theoretical site-directed mutagenesis. Both residues represent highly conserved amino acids when examining various sources of Dihydrofolate Reductase. The first is a phenylalanine located in the center of the pteridine-ring binding site, Phe-31 (*E. coli* numbering scheme; see Figure 3). The modification made and studied was that of a tyrosine mutant, i.e., the addition of a hydroxyl group. The second site for alteration is that of leucine at position 54, forming part of the benz-glutamate portion of the substrate cavity. This mutagenesis work involved both the deletion of the alkyl side chain, i.e., Leu→Gly, and the re-positioning of a methyl substituent, i.e., Leu→Ile.

The Replace facility within BIOGRAF was used to make the mutations to the solvated wild-type protein ternary complex. In all cases, the new side chain was positioned to overlap, as best as possible, with the location of the native side chain. Details of generating the starting structure files are given above. The results of studying these single-site modifications are discussed below.

7.1 Phe-31→Tyr-31

The Phe-31→Tyr-31 mutation involves the addition of a hydroxyl substituent to the aromatic ring of residue 31. This modification introduces a hydrophilic moiety to the pteridine and benzyl binding regions of the active site, thus disturbing the hydrophobic nature of the catalytic pocket. The simulated equilibration, production runs, and structural refinements were handled in the same fashion as those of the wild-type complexes. The results are presented in Tables 34 to 38.

Table 34. Energetics for Tyr-31 mutant DHFR: DHF and NADPH movable, protein ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
1.0	121.60	32.69	99.06	57.93	12.50	7.64	-88.21	1819
2.0	133.70	36.41	98.52	61.09	9.45	3.96	-75.73	1819
3.0	121.54	34.66	100.24	55.42	9.77	3.68	-82.22	1820
4.0	121.77	36.93	98.82	54.08	8.19	5.52	-81.78	1817
AVG	124.65	35.17	99.16	57.13	9.98	5.20	-81.99	1819

DHF-anti Conformer; OD2 protonated.

1.0	109.73	25.32	77.45	82.26	13.31	8.93	-97.54	1877
2.0	120.09	29.26	88.01	81.92	13.72	10.72	-103.53	1883
3.0	122.64	25.38	82.59	92.51	11.93	10.55	-100.32	1867
4.0	117.76	26.06	82.20	91.21	10.41	10.90	-103.03	1867
AVG	117.56	26.51	82.56	86.98	12.34	10.28	-101.11	1874

Table 35. Energetics for Tyr-31 mutant DHFR: DHF movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
1.0	36.89	5.82	15.74	21.47	5.56	7.93	-19.63	548
2.0	43.15	7.05	14.20	25.26	3.93	8.58	-15.88	547
3.0	35.31	5.71	11.13	20.32	4.40	8.77	-15.02	544
4.0	36.16	5.96	11.62	19.58	3.02	10.94	-14.95	545
AVG	37.88	6.14	13.17	21.66	4.23	9.06	-16.37	546

DHF-anti Conformer; OD2 protonated.

1.0	36.89	4.23	19.30	25.86	5.73	11.32	-29.55	556
2.0	43.29	5.22	18.45	30.79	8.21	9.74	-29.13	552
3.0	44.79	4.59	14.81	36.74	5.69	11.55	-28.59	547
4.0	39.10	4.33	12.24	37.42	3.90	11.24	-30.03	547
AVG	41.02	4.59	16.20	32.70	5.88	10.96	-29.33	551

Table 36. Energetics for Tyr-31 mutant DHFR: NADPH movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
1.0	89.78	26.88	83.32	36.46	6.93	2.98	-66.80	810
2.0	95.88	29.36	84.32	35.84	5.52	-1.22	-57.93	810
3.0	89.80	28.95	89.11	35.10	5.38	-2.03	-66.70	817
4.0	89.01	30.97	87.20	34.51	5.17	-2.52	-66.32	813
AVG	91.12	29.04	85.99	35.48	5.75	-0.70	-64.44	813

DHF-anti Conformer; OD2 protonated.

1.0	76.92	21.09	58.15	56.41	7.58	1.03	-67.33	820
2.0	81.97	24.04	69.55	51.13	5.51	5.03	-73.29	827
3.0	82.74	20.79	67.78	55.77	6.23	2.75	-70.59	825
4.0	83.67	21.73	69.96	53.79	6.51	3.57	-71.89	826
AVG	81.33	21.91	66.36	54.28	6.46	3.10	-70.78	825

Table 37. Energetics for Tyr-31 mutant DHFR: DHF, NADPH, and site movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
1.0	-790.98	148.11	415.67	391.75	103.52	-429.52	-1420.52	104693
2.0	-779.02	147.84	414.62	415.11	97.68	-435.07	-1419.20	104742
3.0	-800.61	147.50	422.58	403.47	100.70	-441.70	-1433.15	105498
4.0	-805.21	150.20	420.58	396.98	96.48	-442.23	-1427.21	104814
AVG	-793.96	148.41	418.36	401.83	99.60	-437.13	-1425.02	104937

DHF-anti Conformer; OD2 protonated.

1.0	-803.25	143.10	401.07	422.53	94.77	-455.98	-1408.75	106555
2.0	-773.14	144.29	392.60	441.86	94.53	-445.18	-1401.25	106740
3.0	-767.28	146.40	396.28	442.28	96.94	-445.34	-1403.83	106484
4.0	-750.78	144.91	393.75	453.65	98.72	-442.68	-1399.13	106700
AVG	-773.61	144.68	395.93	440.08	96.24	-447.30	-1403.24	106620

The energies of the Tyr-31 DHF·NADPH complex are of the same magnitude as those of the wild-type protein. However, the OD2 protonated complex is ~7 kcals lower in energy than the corresponding OD1 complex. Most of this difference in stability is due to the effect the mutation has on the energy of the bound NADPH cofactor (see Tables 31 and 36).

The inclusion of the active site protein in the energetic analysis (results given in Table 37) indicate that the OD1 protonated state is energetically favored by ~20 kcal. This preference parallels that found for the wild-type ternary complexes. The energy gap between the preferred OD1 complex and the higher energy OD2 complex has dropped by ~50%. In the wild-type enzyme, the difference between the two DHF-anti structures was ~40 kcals, but in the Tyr-31 mutant the OD1 form is now stabilized by ~20 kcal.

The Tyr-31 mutant's incorporation of a hydroxyl group into the hydrophobic portion of the substrate pocket alters the structural character of the active site and introduces change in the binding orientation of DHF. The hydroxyl group disturbs the active site side chain packing, causing the aromatic ring to swing out toward the bound DHF. This structural change: (a) allows the hydroxyl group to hydrogen-bond to the carbonyl oxygen of Ile-94; (b) forces the side chain of Leu-28 to adjust its equilibrium position; and (c) ultimately alters the conformation of bound DHF. Figure 16 illustrates the changes to substrate and active site side chain orientations produced by the Phe-31→Tyr-31 mutation. A comparison of the active site surfaces mapped out by the wild-type and the Tyr-31 DHFR is shown in Figure 17. The change that is due to the "swinging" of the tyrosine's aromatic ring is apparent in the 3 o'clock position (with respect to the pteridine rings), while the adjustment to Leu-28's alkyl side chain is discernable in the 1-2 o'clock position.

Figure 16. Wire-frame comparison of the relaxed wild-type (red DHF and NADPH, yellow protein) and Tyr-31 mutant (light blue) DHFR ternary complexes. The hydroxyl group of Tyr-31 introduces a change in the binding orientation of the pteridine and benz-glutamate portions of the substate.

Figure 17. The solvent surface comparison of wild-type (red DHF, NADPH, and Phe-31, orange site surface) and Tyr-31 mutant (light blue DHF, NADPH, Tyr-31, blue site surface) DHFR ternary complexes. The change that is due to Tyr-31 is visible to the immediate left of position 31, while the difference produced by Leu-28 is shown northeast of the pteridine ring.

Figure 16.

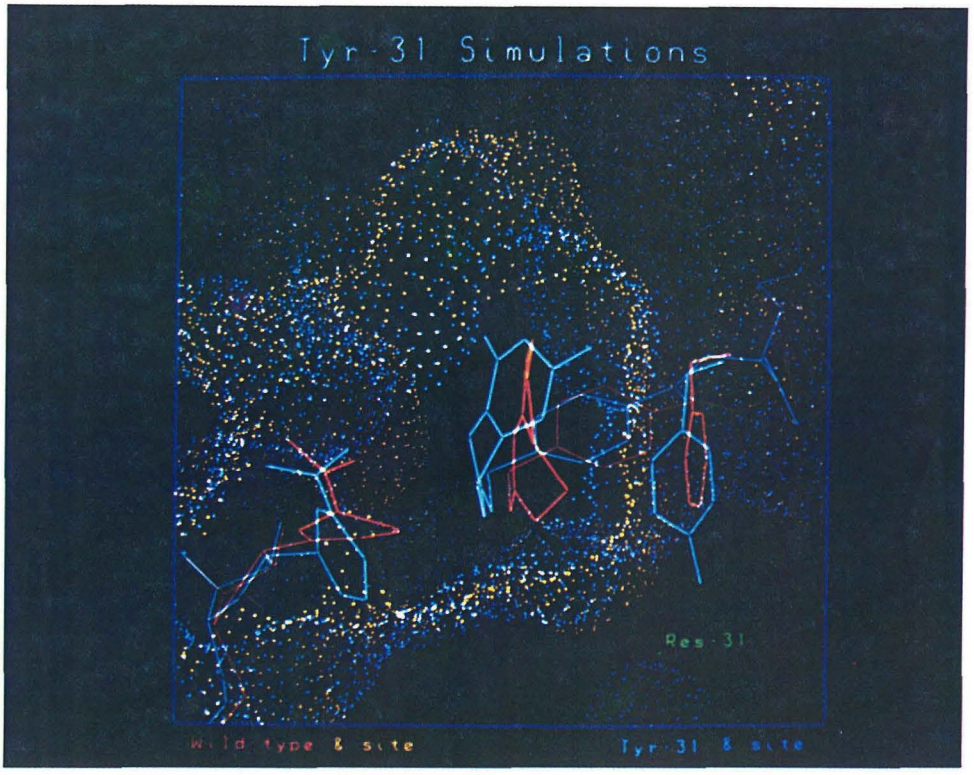


Figure 17.

The hydride transfer distances for the simulated Tyr-31 DHF-anti ternary structures are given in Table 38. The NC4 to C6 distances differ by 1.0 Å for the two states, with the OD2 form having the shorter hydride transfer distance of ~4.0 Å. Thus, the energetically favored Tyr-31 DHF ternary complex leads to a NC4-C6 distance that is ~0.4Å longer than for wild-type, leading to a slower rate for hydride transer.

Experimentally, the Tyr-31 mutant protein has a kinetic profile different from that of the wild-type enzyme.[12] The product off-rate has increased, relative to the wild-type protein; fifty-fold and hydride transfer is now the rate-limiting step in the steady-state cycle. This change in the catalytic profile may be explained by two of the simulation results. First, the increase in product off-rate could be attained by the dissimilarity observed in the substrate binding. Presumably, this change in preferred DHF conformation between the wild-type and Tyr-31 proteins would carry over to a change in the orientation of bound product, thus leading to the measured increase in THF off-rate. Secondly, the slight increase in hydride transfer distance between the Tyr-31 and native forms of DHFR, ~4.0 vs. ~3.6 Å, respectively, should have an effect on the rate for the transfer process, as experimentally observed.

7.2 Leu-54→Ile-54

The Leu-54→Ile-54 mutation involves repositioning of one methyl substituent on the amino acid alkyl side chain. This rearrangement should alter the character of the pteridine binding pocket but not produce any void space requiring the inclusion of additional solvent molecules. The simulated equilibration, production runs, and structural refinements were done as described for the wild-type complexes and the results presented in Tables 39 to 43.

Table 38. Tyr-31 Mutant DHFR: hydride transfer distances.

	DHF-anti	
	OD1	OD2
1.0	4.86	4.02
2.0	4.89	3.84
3.0	5.09	4.06
4.0	5.03	3.97
AVG	4.97	3.97

The energetics of the DHF·NADPH complex indicate that the DHF-anti, OD1 system has the lowest energy (Table 39). The energy of the pair from this Ile-54 state is ~20 kcal lower than the analogous, reactive pair from the wild-type simulations. This decrease is due to an overall lower energy for the DHF in the Ile system (see Table 40). From these DHF-anti, OD1 calculations, DHF's energy is ~23 kcals while the substrate's energy in the native ternary complex is ~44 kcals, a difference on the order of 20 kcals. The energy of NADPH from both forms of the enzyme, with the OD1 protonation and anti form of DHF, is comparable. The energetics for the DHF·NADPH pair from the DHF-anti, OD2 and DHF-xtl, OD1 is ~7-8 kcals higher, while this energy from the DHF-xtl, OD2 complex has dropped by more than 20 kcals for the isoleucine at position 54, relative to the wild-type protein. These differences are due to changes in the individual energies for both the DHF and the NADPH (Tables 40 and 41).

The addition of the active site protein to the energetic analysis indicates that, overall, the effect of moving the methyl group stabilizes the DHF-xtl forms of the ternary complex (Table 42). The active site energies for both of the DHF-anti protonation states agree within ~2 kcals when comparing the Ile-54 with the Leu-54 enzyme. For the DHF-xtl conformations, the situation is quite different. This mutation has stabilized both protonation states, with respect to the wild-type DHFR, by more than 20 kcals. Thus, the relative population of states would be shifted toward the DHF-xtl form for this mutant protein and could alter the steady-state catalytic profile.

The structural modifications at the active site, because of the repositioning of the Leu-54 γ-carbon, are illustrated by Figures 18 and 19. The conformation of Phe-31's aromatic ring has twisted to accommodate the isoleucine side chain, introducing a change in the binding mode of the pteridine ring. Similarly, the

Table 39. Energetics for Ile-54 mutant DHFR: DHF and NADPH movable, protein ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	99.43	31.84	94.64	56.67	10.77	6.71	-101.21	1876
4.0	96.06	34.65	93.04	58.05	9.52	6.90	-106.09	1888
6.0	88.27	30.89	88.77	55.74	9.38	6.61	-103.12	1887
8.0	97.76	30.74	87.94	56.35	8.79	8.26	-94.31	1884
10.0	108.73	35.87	92.21	59.78	9.11	9.67	-97.91	1891
AVG	98.05	32.80	91.32	57.32	9.51	7.63	-100.53	1885

DHF-anti Conformer; OD2 protonated.

2.0	126.45	36.06	96.40	65.86	11.28	8.45	-91.60	1904
4.0	125.27	34.82	103.15	65.89	14.26	7.05	-99.91	1896
6.0	135.81	38.12	101.08	71.44	12.18	8.99	-95.99	1904
8.0	141.14	40.75	106.93	69.61	14.22	7.78	-98.15	1900
10.0	127.59	38.23	95.18	73.21	12.72	7.70	-99.45	1905
AVG	131.25	37.60	100.55	69.20	12.93	7.99	-97.02	1902

DHF-xtl Conformer; OD1 protonated.

2.0	139.69	33.09	91.79	70.73	16.93	8.43	-81.28	1814
4.0	127.80	32.05	89.43	64.73	13.10	9.47	-80.98	1804
6.0	147.90	33.95	96.90	70.96	13.57	11.30	-78.77	1805
8.0	137.36	35.29	90.94	66.46	12.74	10.46	-78.55	1796
10.0	132.80	33.54	90.14	63.79	12.88	10.02	-77.58	1803
AVG	137.11	33.58	91.84	67.33	13.84	9.94	-79.43	1804

DHF-xtl Conformer; OD2 protonated.

2.0	116.46	28.62	90.98	61.46	13.49	10.77	-88.86	1827
4.0	117.66	26.53	92.23	62.35	11.63	14.85	-89.94	1827
6.0	105.23	25.75	85.49	63.04	16.14	11.38	-96.57	1847
8.0	114.23	25.72	92.28	61.87	13.94	15.59	-95.15	1845
10.0	98.41	24.42	85.46	60.19	13.91	12.23	-97.79	1865
AVG	110.40	26.21	89.29	61.78	13.82	12.96	-93.66	1842

Table 40. Energetics for Ile-54 mutant DHFR: DHF movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	22.82	5.21	16.91	14.95	3.91	10.23	-28.38	531
4.0	25.26	6.04	20.57	15.65	3.50	10.41	-30.91	535
6.0	22.03	5.45	18.90	14.46	3.78	10.20	-30.75	535
8.0	20.34	4.89	18.11	13.48	2.82	9.68	-28.63	538
10.0	25.73	6.45	18.66	16.32	3.25	11.35	-30.30	536
AVG	23.24	5.61	18.63	14.97	3.45	10.37	-29.79	535

DHF-anti Conformer; OD2 protonated.

2.0	34.32	8.55	16.59	18.44	2.63	10.36	-22.25	561
4.0	35.65	6.75	15.75	23.05	5.27	9.73	-24.90	557
6.0	38.95	7.14	15.82	22.59	4.68	11.58	-22.86	564
8.0	34.92	7.03	15.30	21.10	5.28	10.37	-24.17	563
10.0	34.24	7.50	14.69	21.30	4.56	10.77	-24.59	561
AVG	35.62	7.39	15.63	21.30	4.48	10.56	-23.75	561

DHF-xtl Conformer; OD1 protonated.

2.0	67.54	10.55	10.96	34.81	10.96	10.77	-10.50	575
4.0	62.17	9.02	12.59	29.97	8.02	11.61	-9.05	574
6.0	74.15	9.54	13.75	36.76	7.77	14.25	-7.92	570
8.0	66.35	10.99	10.91	31.59	7.04	12.88	-7.06	570
10.0	66.33	10.99	12.44	29.55	7.08	12.58	-6.31	573
AVG	67.31	10.22	12.13	32.54	8.17	12.42	-8.17	572

DHF-xtl Conformer; OD2 protonated.

2.0	57.16	9.16	15.07	23.95	8.49	14.64	-14.15	580
4.0	59.65	7.99	13.75	27.13	6.21	15.43	-10.86	577
6.0	53.56	7.68	14.71	24.35	11.11	12.56	-16.85	581
8.0	48.36	6.27	13.24	21.03	8.93	14.76	-15.86	583
10.0	45.77	5.63	14.28	19.95	9.15	14.03	-17.26	583
AVG	52.90	7.35	14.21	23.28	8.78	14.28	-14.99	581

Table 41. Energetics for Ile-54 mutant DHFR: NADPH movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	81.36	26.64	77.73	41.73	6.86	0.78	-72.38	829
4.0	75.84	28.60	72.47	42.39	6.02	0.57	-74.22	834
6.0	71.55	25.44	69.87	41.29	5.60	0.66	-71.31	837
8.0	82.30	25.85	69.83	42.88	5.97	2.53	-64.76	834
10.0	87.92	29.42	73.55	43.47	5.86	2.43	-66.80	833
AVG	79.79	27.19	72.69	42.35	6.06	1.39	-69.89	833

DHF-anti Conformer; OD2 protonated.

2.0	97.88	27.51	79.80	47.42	8.66	1.55	-67.05	823
4.0	96.05	28.07	87.41	42.84	8.99	0.70	-71.96	824
6.0	102.73	30.98	85.26	48.85	7.50	0.78	-70.63	826
8.0	112.24	33.72	91.63	48.51	8.94	0.69	-71.25	821
10.0	99.60	30.73	80.49	51.90	8.16	0.62	-72.29	826
AVG	101.70	30.20	84.92	47.90	8.45	0.87	-70.64	824

DHF-xtl Conformer; OD1 protonated.

2.0	75.04	22.55	80.83	35.92	5.97	-0.81	-69.41	798
4.0	67.79	23.03	76.83	34.75	5.08	-0.84	-71.07	795
6.0	76.26	24.41	83.15	34.20	5.80	-1.48	-69.80	798
8.0	73.37	24.30	80.03	34.87	5.71	-1.10	-70.44	793
10.0	69.12	22.55	77.71	34.24	5.80	-1.17	-70.00	796
AVG	72.32	23.37	79.71	34.80	5.67	-1.08	-70.14	796

DHF-xtl Conformer; OD2 protonated.

2.0	64.05	19.45	75.92	37.51	5.00	-1.34	-72.49	784
4.0	62.70	18.54	78.48	35.23	5.42	1.95	-76.92	790
6.0	57.02	18.07	70.78	38.68	5.03	1.79	-77.33	794
8.0	70.37	19.45	79.03	40.84	5.01	3.41	-77.38	792
10.0	58.61	18.79	71.18	40.24	4.76	1.66	-78.03	800
AVG	62.55	18.86	75.08	38.50	5.04	1.49	-76.43	792

Table 42. Energetics for Ile-54 mutant DHFR: DHF, NADPH, and site movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	-847.82	155.18	427.54	379.96	102.15	-468.24	-1444.42	105331
4.0	-845.50	158.56	424.66	376.14	98.72	-459.16	-1444.42	105475
6.0	-846.90	152.64	431.54	358.60	93.38	-464.12	-1418.94	105929
8.0	-844.77	151.09	420.98	355.67	93.33	-469.14	-1396.70	106032
10.0	-804.02	158.75	417.27	368.25	91.88	-449.95	-1390.23	105880
AVG	-837.80	155.24	424.40	367.72	95.89	-462.12	-1418.94	105729

DHF-anti Conformer; OD2 protonated.

2.0	-830.46	151.85	414.74	358.55	87.63	-443.03	-1400.19	105660
4.0	-804.21	152.09	406.36	389.72	97.22	-426.48	-1423.12	105377
6.0	-771.73	153.14	419.00	412.81	97.46	-432.24	-1421.89	105768
8.0	-805.75	151.17	409.01	402.22	97.96	-428.28	-1437.81	105748
10.0	-786.64	157.60	397.22	412.36	95.90	-424.39	-1425.34	105572
AVG	-799.76	153.17	409.27	395.13	95.23	-430.88	-1421.67	105625

DHF-xtl Conformer; OD1 protonated.

2.0	-859.10	162.46	394.27	369.34	93.47	-436.95	-1441.69	105595
4.0	-877.89	158.85	381.03	373.46	92.89	-445.01	-1439.12	105840
6.0	-887.96	160.15	381.58	378.46	97.45	-435.92	-1469.68	105796
8.0	-901.77	162.34	375.64	373.09	97.84	-442.54	-1468.14	105588
10.0	-902.62	159.69	379.49	370.69	95.42	-446.79	-1461.13	106006
AVG	-885.87	160.70	382.40	373.01	95.41	-441.44	-1455.95	105765

DHF-xtl Conformer; OD2 protonated.

2.0	-854.04	152.07	376.90	373.28	92.49	-423.28	-1425.50	103510
4.0	-837.71	151.91	384.96	369.13	89.66	-422.21	-1411.17	103876
6.0	-820.33	155.61	390.43	374.56	104.61	-421.19	-1424.35	103901
8.0	-815.76	161.62	392.70	365.00	102.32	-419.37	-1418.03	104527
10.0	-844.30	158.05	389.71	341.20	96.26	-414.43	-1415.10	104443
AVG	-834.43	155.85	386.94	364.63	97.07	-420.10	-1418.83	104051

Figure 18. The equilibrated wild-type DHFR active site region with DHF-anti and NADPH bound (shown in red). Leu-54 (yellow) is shown to the right of the benz-glutamate portion of DHF, while Phe-31 is visible above the pteridine ring.

Figure 19. The equilibrated Ile-54 mutant DHFR active site region with DHF-anti and NADPH bound (shown in green). Ile-54 (white) is shown to the right of the benz-glutamate portion of DHF, while Phe-31, slightly twisted relative to its position in the wild-type protein, is visible above the pteridine ring.

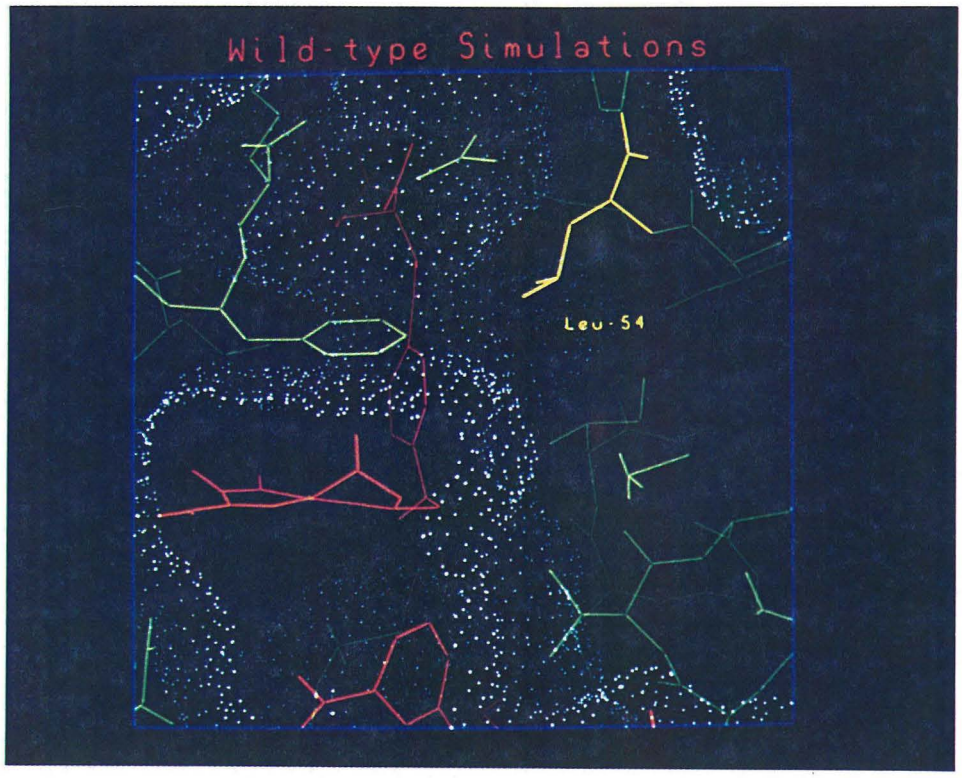


Figure 18.

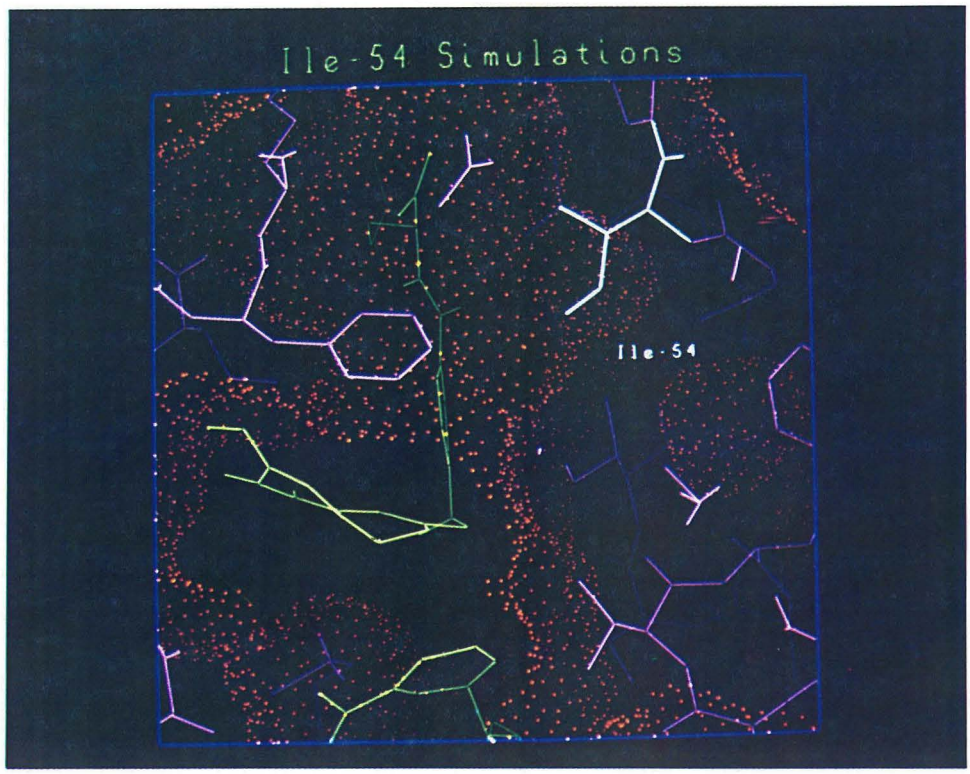


Figure 19.

substrate's benzyl ring position has been slightly twisted. The side chains of Met-42 and Ile-94 modify their positions in response to the mutation.

The Ile substitution at position 54 produces changes in the hydride transfer distance between the nicotinamide and pteridine rings. The distance data are listed in Table 43. Both DHF-anti complexes have comparable distances for the Ile series, a carbon-carbon separation of ~3.9 Å. This hydride transfer value is a bit longer than the 3.6 Å calculated for the native enzyme complexes. The DHF-xtl species, however, have NC4 to C6 separations of ~5.4 Å (OD1) and 4.8 Å (OD2), values significantly longer than those calculated for the Anti forms.

Experimentally, the steady-state kinetic profile for the Ile-54 mutant protein is different from that of the wild-type enzyme.[14] The hydride transfer rate is decreased by a factor of 30 and now competes for the rate-limiting step. This change in the catalytic profile can be understood on the basis of the increase in hydride transfer distance relative to the wild-type protein. As presented above, the transfer distance is ~3.6 Å for the native protein while the value for the Ile-54 case is ~3.9 Å. This slight change in transfer distance could be significant enough to account for the changes found experimentally.

7.3 Leu-54→Gly-54

The Leu-54→Gly-54 mutation involves removal of the leucine alkyl side chain. Such a deletion creates an empty region in the active site of the ternary complex. To fill this void cavity, two additional water molecules were substituted in each of the four starting structure files. The simulated equilibration, production runs, and structural refinements were handled in the same fashion as those of the wild-type complexes. The results are presented in Tables 44 to 48.

The energetics of the Gly-54 DHF·NADPH complex is quite different from

Table 43. Ile-54 Mutant DHFR: hydride transfer distances.

	DHF-anti		DHF-xtl	
	OD1	OD2	OD1	OD2
2.0	4.06	3.79	5.36	4.86
4.0	3.93	3.79	5.47	5.05
6.0	3.87	3.87	5.28	4.74
8.0	3.86	3.77	5.40	4.91
10.0	4.03	3.90	5.36	4.31
AVG	3.95	3.82	5.37	4.77

that found in the wild-type protein (Tables 44 and 29, respectively). The set from the DHF-xtl with OD2 protonated complex is the most stable form for the Gly-54 series, in contrast to the DHF-anti OD1 set for the native protein. The energies of the DHF-anti OD2 and both DHF conformers with OD1 protonated are ~20 kcals higher in the Gly-54 set than they are in the Leu-54 studies. These differences are attributable to the energy changes in the individual molecules (see Tables 45 and 46): (a) the 15-20 kcal increase in energy of the DHF for both OD1 species and (b) the 20 kcal change in NADPH from the DHF-anti OD2 complex.

Including the active site protein in the energetic profile maintains the overall stability of the DHF-xtl OD2 Gly-54 ternary complex (data given in Table 47). This form is ~30 kcal lower in energy than either of the OD1 protonated complexes but only 16 kcal different from the DHF-anti OD2 state. Direct energetic comparisons of the Gly-54 series with the native set are not possible within the molecular mechanics framework because the two enzymes are composed of a different number of atoms.

The structural changes resulting from the removal of the Leu-54 side chain are illustrated by Figure 20. The conformation of Phe-31's aromatic ring has twisted to a much greater extent than in the Ile-54 structures, introducing a change in the binding of DHF's pteridine ring. The benzyl ring of the substrate has also been twisted slightly. The conformation of Ile-50's alkyl side chain exhibits the most pronounced change of any active site residue. The side chain re-orients itself, to substitute partially for the absent Leu group, by rotating its δ_1 carbon into the vacant site. This conformational change results in a ~3.5Åchange of position for the δ_1 carbon of Ile-50! Less dramatic shifts are found for the side chain positions of Met-42 and Ile-94, similar to their changes in the Ile-54 mutant DHFR complexes.

Table 44. Energetics for Gly-54 mutant DHFR: DHF and NADPH movable, protein ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	131.31	36.34	85.20	83.57	16.03	1.44	-91.27	1861
4.0	136.12	37.78	85.67	87.64	14.25	3.21	-92.44	1864
6.0	141.32	40.42	87.65	87.12	15.09	4.61	-93.56	1848
8.0	147.14	42.02	92.54	87.03	15.63	4.82	-94.92	1848
10.0	148.66	45.73	92.14	88.88	14.45	2.30	-94.84	1851
AVG	140.91	40.46	88.64	86.85	15.09	3.28	-93.41	1854

DHF-anti Conformer; OD2 protonated.

2.0	131.35	42.58	109.01	58.20	9.70	5.40	-93.54	1829
4.0	133.28	41.72	111.08	62.52	11.54	5.88	-99.46	1824
6.0	140.02	46.80	119.88	57.30	11.23	3.88	-99.07	1828
8.0	137.66	43.78	118.50	59.34	12.18	4.73	-100.86	1823
10.0	155.65	49.38	123.88	63.46	13.34	5.57	-99.97	1821
AVG	139.59	44.85	116.47	60.16	11.60	5.09	-98.58	1825

DHF-xtl Conformer; OD1 protonated.

2.0	158.35	39.59	98.18	84.66	12.96	9.03	-86.07	1891
4.0	145.71	33.43	92.61	83.59	12.23	12.96	-89.11	1872
6.0	159.01	36.51	97.70	87.70	13.95	12.89	-89.73	1878
8.0	142.37	33.96	88.77	84.52	13.15	10.68	-88.72	1862
10.0	148.29	34.08	92.74	82.63	13.90	13.33	-88.39	1863
AVG	150.75	35.51	94.00	84.62	13.24	11.78	-88.40	1873

DHF-xtl Conformer; OD2 protonated.

2.0	116.77	29.35	91.22	74.88	11.50	6.16	-96.34	1927
4.0	124.27	29.90	84.82	79.84	12.16	8.52	-90.96	1940
6.0	113.35	27.48	87.99	68.77	11.72	6.26	-88.87	1923
8.0	122.97	32.84	84.52	75.54	10.05	8.07	-88.04	1940
10.0	116.99	29.41	85.44	69.95	10.69	8.01	-86.51	1932
AVG	118.87	29.80	86.80	73.80	11.22	7.40	-90.14	1932

Table 45. Energetics for Gly-54 mutant DHFR: DHF movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	55.48	8.23	17.37	42.93	11.17	9.72	-33.94	529
4.0	55.14	6.57	16.50	45.74	10.05	9.84	-33.55	531
6.0	60.17	9.00	19.74	45.29	10.22	12.12	-36.20	527
8.0	64.60	9.97	18.09	47.50	10.89	11.41	-33.26	524
10.0	62.64	9.41	18.51	48.90	10.20	9.93	-34.30	530
AVG	59.61	8.64	18.04	46.07	10.51	10.60	-34.25	528

DHF-anti Conformer; OD2 protonated.

2.0	31.89	6.14	11.80	28.03	3.51	10.73	-28.33	535
4.0	36.50	7.97	14.33	30.96	4.37	10.48	-31.61	535
6.0	32.55	8.27	13.91	26.95	3.46	9.39	-29.42	5.34
8.0	35.26	7.86	14.90	28.23	4.53	9.35	-29.62	536
10.0	39.33	7.49	14.25	32.11	5.37	9.80	-29.70	535
AVG	35.11	7.55	13.84	29.26	4.25	9.95	-29.74	535

DHF-xtl Conformer; OD1 protonated.

2.0	63.54	6.76	20.57	35.39	5.79	11.35	-16.32	553
4.0	71.00	7.95	20.39	37.53	6.31	12.43	-13.61	552
6.0	76.44	8.03	22.27	40.20	8.07	13.46	-15.59	555
8.0	71.52	7.43	20.76	36.93	7.09	12.47	-13.15	549
10.0	70.60	6.77	21.61	35.21	7.51	12.62	-13.12	552
AVG	70.62	7.39	21.12	37.05	6.95	12.47	-14.36	552

DHF-xtl Conformer; OD2 protonated.

2.0	42.92	7.35	13.05	24.65	4.16	8.70	-14.99	569
4.0	53.04	6.89	14.10	30.04	4.72	10.02	-12.73	571
6.0	46.20	6.25	14.54	23.03	4.58	9.96	-12.16	572
8.0	51.39	6.09	14.36	28.73	3.72	10.69	-12.20	566
10.0	50.34	7.42	15.57	23.92	4.15	10.14	-10.87	568
AVG	48.78	6.80	14.32	26.07	4.27	9.90	-12.59	569

Table 46. Energetics for Gly-54 mutant DHFR: NADPH movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	83.05	28.11	67.82	40.64	4.86	-3.36	-55.02	808
4.0	88.32	31.21	69.18	41.90	4.20	-1.75	-56.43	808
6.0	87.70	31.42	67.91	41.83	4.87	-3.55	-54.77	805
8.0	88.73	32.05	74.46	39.53	4.74	-2.71	-59.34	798
10.0	92.40	36.32	73.63	39.98	4.25	-3.53	-58.25	802
AVG	88.04	31.82	70.60	40.78	4.58	-2.98	-56.76	804

DHF-anti Conformer; OD2 protonated.

2.0	103.92	36.44	97.21	30.16	6.19	-1.76	-64.32	797
4.0	100.80	33.75	96.75	31.56	7.17	-1.02	-67.41	794
6.0	112.69	38.54	105.96	30.35	7.78	-1.44	-68.49	791
8.0	106.93	35.92	103.60	31.11	7.65	-1.12	-70.22	790
10.0	120.62	41.88	109.62	31.34	7.97	-0.99	-69.22	790
AVG	108.99	37.31	102.63	30.90	7.35	-1.27	-67.93	792

DHF-xtl Conformer; OD1 protonated.

2.0	103.00	32.83	77.61	49.27	7.17	2.42	-66.30	811
4.0	82.76	25.48	72.22	46.06	5.92	4.87	-71.79	817
6.0	90.07	28.48	75.43	47.50	5.88	3.82	-71.03	820
8.0	78.59	26.53	68.02	47.60	6.06	2.67	-72.28	818
10.0	84.65	27.30	71.13	47.43	6.39	4.66	-72.26	818
AVG	87.81	28.12	72.88	47.57	6.28	3.69	-70.73	817

DHF-xtl Conformer; OD2 protonated.

2.0	81.27	22.00	78.17	50.23	7.34	0.88	-77.36	824
4.0	78.01	23.01	70.72	49.81	7.44	3.12	-76.08	827
6.0	73.58	21.23	73.46	45.74	7.14	0.56	-74.54	821
8.0	78.71	26.75	70.15	46.81	6.32	1.47	-72.80	828
10.0	73.37	21.98	69.86	46.03	6.55	1.85	-72.90	828
AVG	76.99	22.99	72.47	47.72	6.96	1.58	-74.74	826

Table 47. Energetics for Gly-54 mutant DHFR: DHF, NADPH, and site movable, all else ignored.

DHF-anti Conformer; OD1 protonated.

Time	E_{Tot}	E_b	E_θ	E_ϕ	E_ω	E_{vdW}	E_{el}	NBs
2.0	-819.23	153.31	400.25	397.15	111.34	-445.56	-1435.72	104292
4.0	-783.71	154.70	410.49	416.36	104.93	-433.91	-1436.28	104227
6.0	-758.90	161.02	416.63	423.38	108.43	-429.77	-1438.59	104115
8.0	-724.10	165.58	423.64	437.29	111.72	-424.10	-1438.24	103723
10.0	-785.17	165.61	416.08	416.24	107.69	-433.80	-1456.99	103840
AVG	-774.22	160.04	413.42	418.08	108.82	-433.43	-1441.16	104039

DHF-anti Conformer; OD2 protonated.

2.0	-829.44	166.13	433.78	403.58	95.08	-427.78	-1500.23	104388
4.0	-796.80	162.25	428.75	427.24	98.88	-437.52	-1476.39	104460
6.0	-801.45	169.43	430.91	412.79	97.31	-441.84	-1470.06	104688
8.0	-759.34	168.23	444.10	427.41	108.65	-432.68	-1475.04	104335
10.0	-750.65	177.13	449.58	438.59	111.29	-437.68	-1489.56	104298
AVG	-787.54	168.63	437.52	421.92	102.24	-435.50	-1482.26	104434

DHF-xtl Conformer; OD1 protonated.

2.0	-765.84	162.84	399.54	429.03	105.10	-441.01	-1421.34	105414
4.0	-809.59	155.50	394.39	404.00	98.57	-434.41	-1427.63	106194
6.0	-779.65	166.95	398.27	423.97	105.05	-433.36	-1440.53	106484
8.0	-754.34	163.30	399.17	437.11	109.25	-439.11	-1424.06	106530
10.0	-751.83	163.39	409.23	418.07	107.16	-423.86	-1425.83	106603
AVG	-772.25	162.40	400.12	422.44	105.03	-434.35	-1427.88	106245

DHF-xtl Conformer; OD2 protonated.

2.0	-819.65	147.37	387.62	434.44	109.67	-442.58	-1456.19	105608
4.0	-799.45	152.12	379.37	446.69	113.45	-431.03	-1460.04	105627
6.0	-790.73	149.97	380.62	445.02	119.83	-430.32	-1455.85	105492
8.0	-799.26	155.33	372.27	446.76	114.41	-435.26	-1452.76	105.771
10.0	-808.51	148.18	374.82	438.92	115.34	-438.54	-1447.23	105460
AVG	-803.52	150.59	378.94	442.37	114.54	-435.55	-1454.41	105592

Figure 20. Equilibrated Gly-54 mutant DHFR active site region with DHF-anti and NADPH bound (shown in blue). Gly-54 (red) is shown to the right of the benz-glutamate portion of DHF, while Phe-31, slightly twisted relative to its position in the wild-type protein, is visible above the pteridine ring. The water molecules substituting for the removed side chain give rise to the "new" surface dots that appear between the residue 54 site and the DHF.

Figure 21. A comparison of the substrate and cofactor from the equilibrated ternary complexes used for the Res-54 studies.

Figure 20.

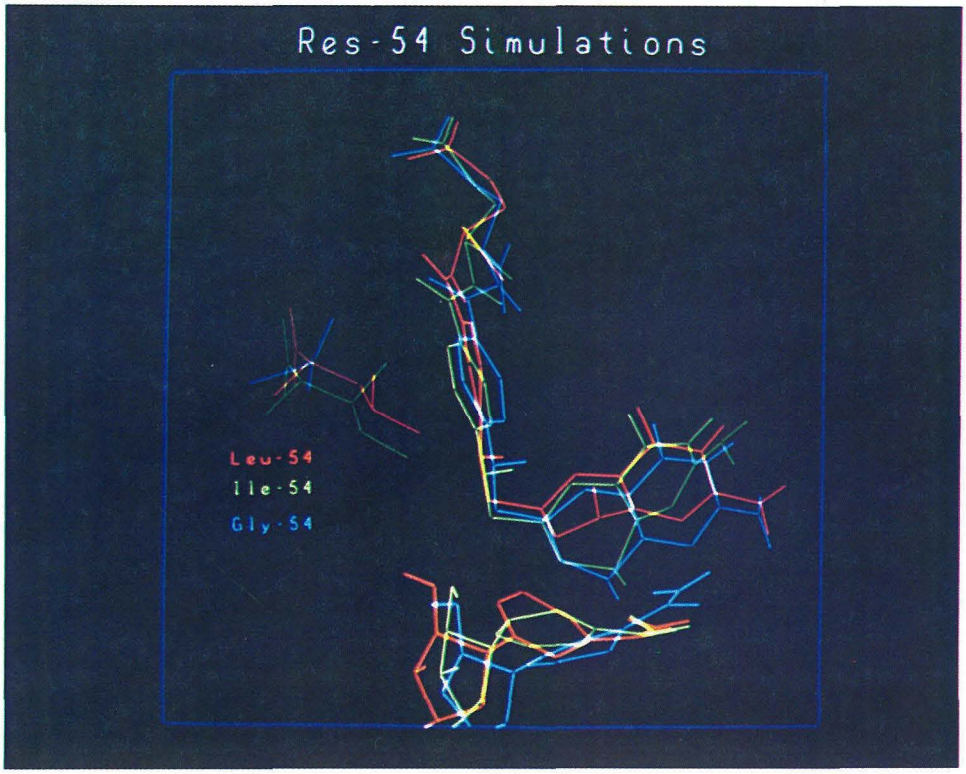


Figure 21.

The hydride transfer distances for the simulated Gly-54 complexes are given in Table 48. The values for each of the four states are approximately equivalent; i.e., they are within 0.2 Å of each other. Thus, for the Gly-54 DHFR enzyme, the calculated ground-state, hydride transfer distance is ~4.4 Å.

Figure 21 shows a comparison of the equilibrated DHF·NADPH orientations from simulations of Res-54 mutants.

Experimentally, the Gly-54 mutant protein has a kinetic profile different from that of the wild-type enzyme.[9] The rate of hydride transfer from the nicotinamide ring to C6 of the pteridine has dropped enough now to be the rate-limiting step in the steady-state cycle. This change in the catalytic profile may be explained on the basis of the difference in hydride transfer distance between the Gly-54 and native forms of DHFR. As indicated above, the simulated transfer distance in the wild-type protein is ~3.6 Å, while the corresponding value for the Gly-54 case is ~4.4 Å, an increase of 0.8 Å! Clearly, that relatively large increase should have a substantial effect on the rate for the transfer process, as was experimentally observed.

Another interesting insight comes from these Gly-54 calculations. For the *wild-type* protein, theory finds that the low energy and short hydride transfer distance both correspond to a DHF-anti conformation of bound substrate. This is in agreement with the isotope labeling studies, which showed that the reactive DHF does indeed have its pteridine ring in a conformation different from that of MTX when bound to DHFR. For the Gly-54 enzyme, however, simulation suggests that the substrate binds and reacts in a conformation similar to that of MTX. This conformation of DHF would yield THF product with the alternate stereochemistry at C6. Isotope studies on the Gly-54 produced THF will give a definitive indication of the nature of this product and hence will test these

Table 48. Gly-54 Mutant DHFR: hydride transfer distances.

	DHF-anti		DHF-xtl	
	OD1	OD2	OD1	OD2
2.0	4.47	4.29	4.26	4.29
4.0	4.33	4.28	4.61	4.25
6.0	4.42	4.38	4.43	4.44
8.0	4.53	4.27	4.48	4.38
10.0	4.48	4.25	4.67	4.18
AVG	4.45	4.29	4.49	4.31

models of the DHFR ternary complex.

The simulated ground-state hydride transfer distances for the mutant DHFR proteins investigated are summarized in Table 49.

Table 49. Summary of simulated ground-state hydride transfer distance for DHFR·DHF·NADPH ternary complexes.

	DHF-anti		DHF-xtl	
	OD1	OD2	OD1	OD2
Wild-type	3.6	4.4	4.5	5.6
Ile-54	4.0	3.8	5.4	4.8
Gly-54	4.5	4.3	4.5	4.3
Tyr-31	5.0	4.0		

8. Segment Mutations

The extension of the structure-sequence homology studies to include *Chicken* DHFR brought to light the possibility of modifying the bacterial enzyme in such a way as to incorporate some of the more typical mammalian properties, i.e., the ability to readily reduce folate (FOL) to 7,8-dihydrofolate. Bacterial forms of Dihydrofolate Reductase catalyze the transformation of FOL to DHF very slowly, while the *Chicken* enzyme more readily facilitates the reduction.[19] Molecular modeling/docking studies, discussed in Section 3, suggest that the salient difference between bacterial and mammalian ease of FOL reduction is directly related to the size of the substrate binding pocket. It is possible that this explains the difference in chemistry, because a larger and/or more flexible catalytic pocket could accommodate the pteridine ring conformation suitable for hydride transfer to the C7 position and thus catalyze the reduction of folate.

The docking and comparison studies, discussed in detail in Section 3, illustrated two regions of dissimilarity between the *E. coli* and *Chicken* forms of DHFR that were appropriate segments to target for this protein-engineering effort. These sites are illustrated by Figure 22. Both regions involve sequence insertions in segments that are adjacent to the substrate binding pocket, as shown in Figure 23. The first targeted segment mutation incorporates the transformation of the *E. coli* residues Gly-51 through Asp-59 to the corresponding *Chicken* loop Pro-61 through Asn-72. In addition to altering the amino acid sequence, this mutation also increases the size of this loop region by three amino acid insertions. The second modification involves substituting the nine amino acid loop Ser-39 to Gln-47 from the *Chicken* for Leu-36 *E. coli*.

The molecular simulations used to investigate these modified proteins focused on the ability of the protein to facilitate folate reduction. To this end,

Figure 22. The docked *E. coli* (yellow protein backbone and red MTX) and *Chicken* (magenta protein backbone and green NADPH) DHFR crystal complexes depicting the agreement in overall protein fold.

Figure 23. The active site region of the docked *E. coli* and *Chicken* complexes illustrating the two *Chicken*-loop regions identified for use in the hybrid mutation studies.

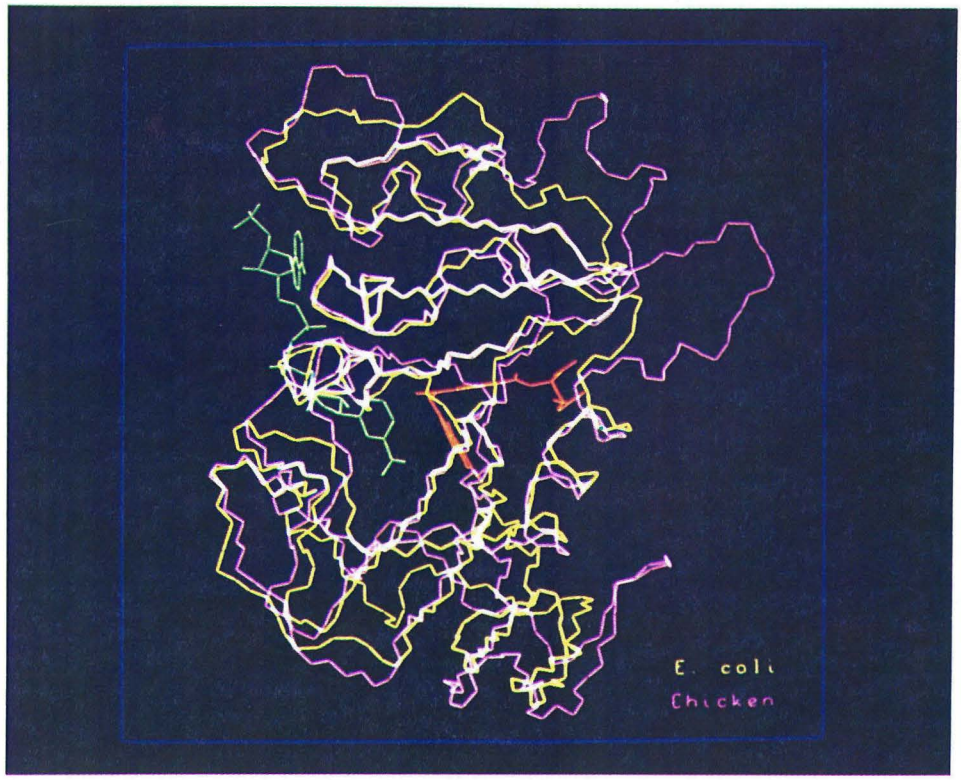


Figure 22.

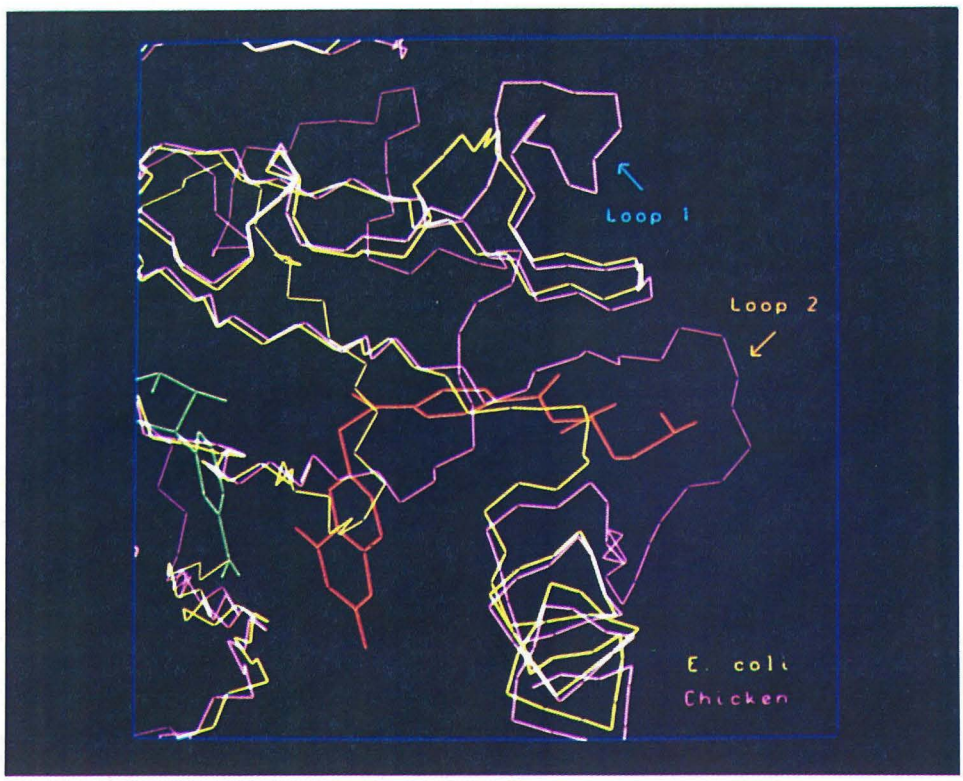


Figure 23.

simulation strategies were designed to permit insight on the structural and energetic stability of folate ternary complexes. The "assay" used to decide whether the enzymes should accommodate, and hence reduce, folate involves comparing the results from constrained and unconstrained simulation runs. The constraints employed for the calculations hold the substrate (either folate or dihydrofolate) and NADPH in a position suitable for reactivity.

The protocols used in these loop hybrid investigations involved:

(1) Generation of initial hybrid protein structures based upon the crystallographic structures of the *E. coli* and *Chicken* forms of DHFR. The coordinates from the docked *Chicken* structure were used for the substituted amino acid segments.

(2) Completion of the starting structure files by adding substrate (either DHF or FOL), NADPH, and adequate solvent.

(3) Incorporation of structural constraints between the substrate and cofactor appropriate for catalytic reduction, if required by the calculations.

(4) Relaxation of the *full* ternary complex via molecular dynamics simulations.

(5) Comparison of the structures and energetic profiles from the corresponding constrained and nonconstrained simulations.

Because of the nature and extent of the structural reorganization expected for these protein hybrids and folate ternary complexes, extended molecular dynamics simulations are required. Such studies might require a few months of elapsed time. Preliminary simulation runs were undertaken to determine: (a) if the simulation strategies are appropriate and (b) if the *full* theoretical component of the project should be undertaken (post-PhD).

The calculations were carried out on both loop substitution proteins as well as the native *E. coli* enzyme. The preliminary simulations, which include ini-

tial structure relaxation, and only 2 ps of full production equilibration, indicate subtle structural differences between the constrained and nonconstrained protein ternary complexes, as given in Table 50.

The RMS deviations for the wild-type *E. coli* DHF ternary complexes are ~0.3, with all heavy atoms and only protein backbone atoms. The corresponding deviations for the FOL complexes are ~0.4, slightly higher than the DHF values. This difference may be due to the protein's catalytic properties. Its ability to reduce dihydrofolate requires the accommodation of bound DHF and NADPH in a reactive orientation. This enzyme's difficulty in transforming folate, however, suggests that an unfavorable situation may occur when FOL and NADPH are brought together in a reactive fashion. The difference in RMS deviations for the *E. coli* DHF and FOL ternary complexes support this hypothesis.

The RMS comparisons indicate a decrease in the dissimilarity of the constrained versus nonconstrained complexes for the hybrid proteins. Loop hybrid 1, with its addition of 3 amino acids to the substrate binding pocket, exhibits RMS values for both DHF and FOL ternary complexes, slightly smaller than those of the wild-type protein. This decrease is even more pronounced for loop hybrid 2; the DHF RMS value is more than 0.1 lower, while the FOL deviation has dropped by ~0.2. The incorporation of eight additional residues at the back of the benz-glutamate pocket may be responsible for this improvement in structural integrity.

Visual comparisons of each of the hybrid FOL ternary complexes (including constraints) with the *E. coli* wild-type FOL ternary complex are shown in Figures 24 and 25.

The energetic comparisons from the hybrid protein simulations (after only 2 ps) were inconclusive. The differences between the energies for the constrained

Table 50. RMS deviations of constrained and nonconstrained structures for full DHFR ternary complexes.

	Heavy	Main
	Wild-type *E. coli*	
DHF	0.31	0.32
FOL	0.42	0.36
	Hybrid 1	
DHF	0.26	0.28
FOL	0.36	0.35
	Hybrid 2	
DHF	0.19	0.16
FOL	0.23	0.21

Figure 24. Preliminary results from the reactive folate simulations on loop hybrid 1 (shown in blue) and wild-type *E. coli* (drawn in red).

Figure 25. Preliminary results from the reactive folate simulations on loop hybrid 2 (shown in green) and wild-type *E. coli* (drawn in red).

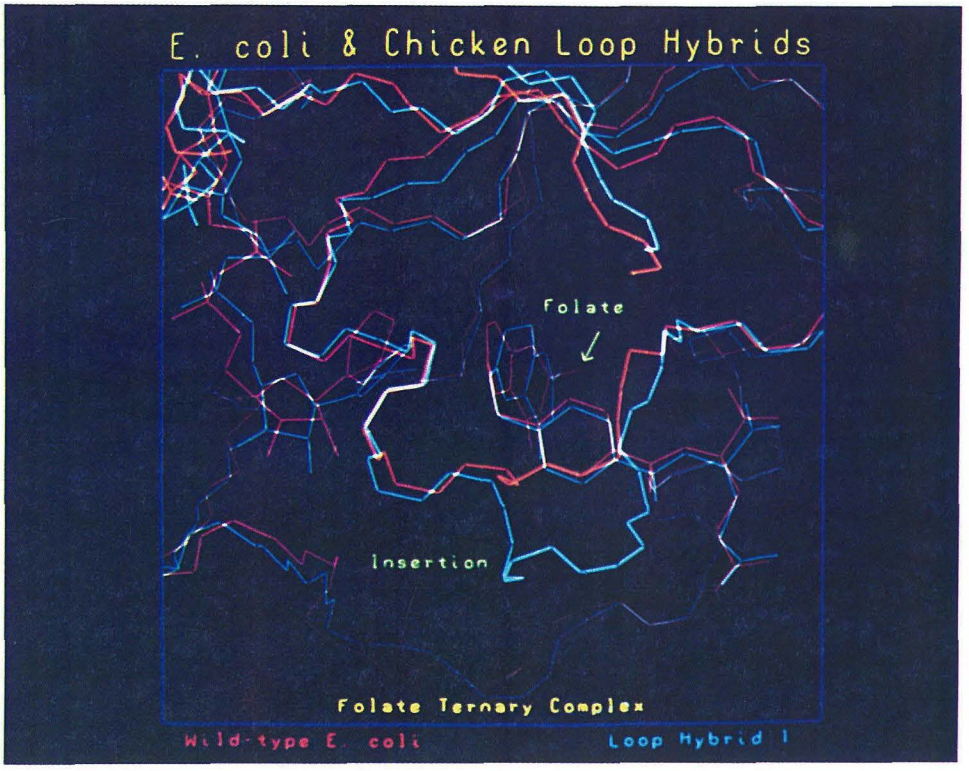


Figure 24.

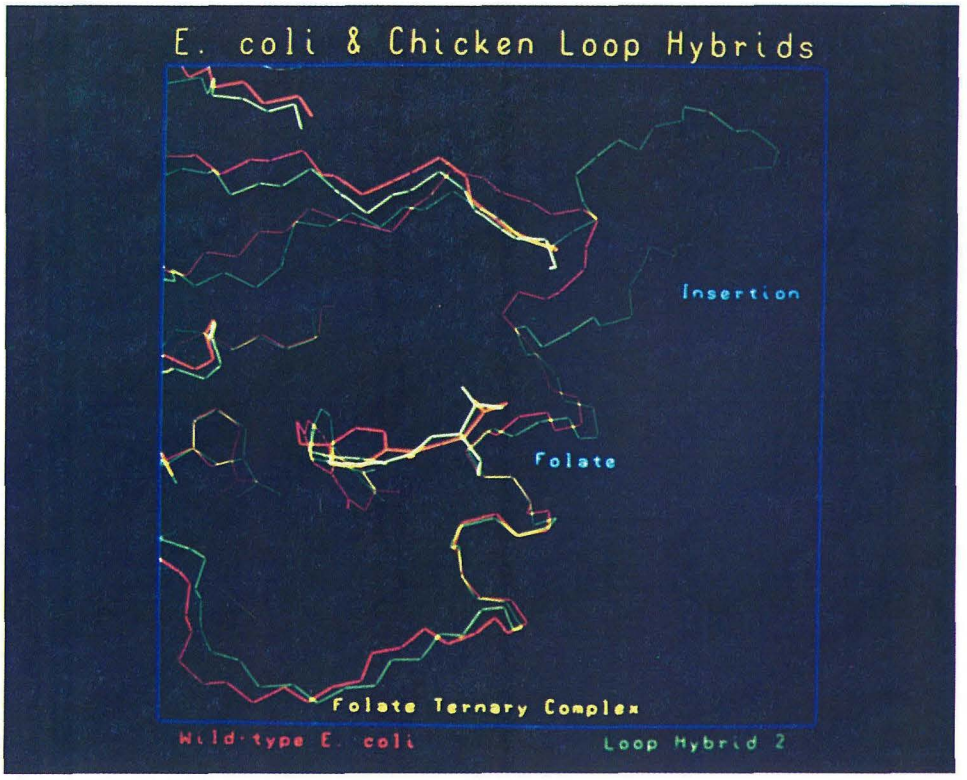


Figure 25.

and the nonconstrained complexes were negligible. Additional simulation should allow for a differentiation of the constrained and nonconstrained energetic profiles from the wild-type, loop hybrid 1, and loop hybrid 2 DHF and FOL ternary complexes.

Thus, these preliminary molecular dynamics simulations on the modified *E. coli*·*Chicken* loop hybrid materials indicate that the increase in substrate binding pocket allows for a higher degree of structural integrity when maintaining a reactive *folate* ternary complex. A more extensive series of theoretical investigations will be undertaken to further probe the structural and folate-reducing properties of these hybrid enzymes.

9. Summary

Investigations of the similarity in sequence, structure, and chemical homology of the *E. coli* and *L. casei* forms of DHFR, using computer graphics with macromolecular docking in conjunction with chemical homology assignments, offer an explanation for the agreement in catalytic reactivity of these bacterial enzymes. The overall sequence homology between the *E. coli* and *L. casei* proteins is ~27%, while the sequence homology for the active site region is ~35%. The *chemical homology* for the entire catalytic pocket, however, is ~70%! The extension of this structure-sequence study to include *Chicken* DHFR indicates only ~60% chemical homology with the bacterial forms, thus suggesting subtle differences in the character of the active site cavity. These sequence-structure comparisons between bacterial and *Chicken* DHFR also identified target sites for engineering *E. coli* · *Chicken* hybrid enzymes.

The molecular simulations on the native form of *E. coli* Dihydrofolate Reductase indicate: (a) the energetically favored form of methotrexate's pteridine ring, when bound to the enzyme, is that found in the crystal studies; (b) the active site region will adjust to accommodate the xtl and the anti forms of both MTX and the natural substrate, dihydrofolate; and (c) the anti-DHF (with Asp-27 OD1 protonated) ternary complex is the catalytically favored form with hydride transfer distance of ~3.6 Å.

The computational investigations on single-site mutants reveal structural insights that offer explanations for trends in the experimental kinetic data. Phe-31→Tyr-31 introduces change in substrate, and presumably product, binding, thus increasing the product off-rate. The removal of the alkyl side chain associated with the Leu-54→Gly-54 mutation produces a dramatic increase in separation between the nicotinamide and pteridine reaction centers (from ~3.6 Å

to ~4.3 Å), which parallels the mutant protein's increase in hydride transfer rate. Similarly, the repositioning of a methyl group (Leu-54 → Ile-54) alters the hydride transfer distance and rate, as summarized in Table 49.

Preliminary molecular simulations on the engineered *E. coli·Chicken* hybrid materials suggest that the increase in the substrate binding pocket, resulting from the loop insertions, allows for a higher degree of structural integrity when maintaining a reactive *folate* ternary complex.

10. Conclusions

The combination of theory and experiment on the mechanistic properties of wild-type and mutant complexes of Dihydrofolate Reductase provides complimentary information that together has led to new insights and understanding not accessible on the basis of either alone. As the predictions on these systems are tested via experimentation, and as the reliability of the predictions increases, these techniques will become true design tools for the development of new biological systems.

Notes and References

(1.) J. E. Gready. *Adv. Pharm. Chemotherapy* **1980**, *17*, 37.

(2.) L. F. Kuyper, B. Roth, D. P. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. Norrington, D. J. Baker, P. J. Goodford. *J. Med. Chem.* **1985**, *28*, 303.

(3.) D. A. Matthews, R. A. Alden, J. T. Bolin, D. J. Filman, S. T. Freer, R. Hamlin, W. G. J. Hol, R. L. Kisliuk, E. J. Pastore, L. T. Plante, N.-H. Xuong, J. Kraut. *J. Biol. Chem.* **1978**, *253*, 6946.

(4.) D. A. Matthews, R. A. Alden, S. T. Freer, N.-H. Xuong, J. Kraut. *J. Biol. Chem.* **1979**, *254*, 4144.

(5.) J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin, J. Kraut. *J. Biol. Chem.* **1982**, *257*, 13650.

(6.) D. J. Filman, J. T. Bolin, D. A. Matthews, J. Kraut. *J. Biol. Chem.* **1982**, *257*, 13663.

(7.) D. R. Smith, J. M. Calvo. *Nuc. Acids Res.* **1980**, *8*, 2255.

(8.) J.-T. Chen, R. J. Mayer, C. A. Fierke, S. J. Benkovic. *J. Cell Biochem.* **1985**, *29*, 73.

(9.) R. J. Mayer, J.-T. Chen, K. Taira, C. A. Fierke, S. J. Benkovic. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7718.

(10.) E. E. Howell, J. E. Villafranca, M. S. Warren, S. J. Oatley, J. Kraut. *Science* **1986**, *231*, 1123.

(11.) C. A. Fierke, K. A. Johnson, S. J. Benkovic. *Biochemistry* **1987**, *26*, 4085.

(12.) J. T. Chen, K. Taira, C. D. Tu, S. J. Benkovic. *Biochemistry* **1987**, *26*, 4093.

(13.) S. J. Benkovic, C. A. Fierke, A. M. Naylor. *Science* **1988**, *239*, 1105.

(14.) D. J. Murphy, S. J. Benkovic. *Biochemistry* **1989**, in press.

(15.) S. J. Benkovic, J. A. Adams, C. A. Fierke, A. M. Naylor. *Pteridines* **1989**, in press.

(16.) P. A. Charlton, D. W. Young, B. Birdsall, J. Feeney, G. C. K. Roberts. *J. Chem. Soc., Chem. Commun.* **1979**, 922.

(17.) S. R. Stone, J. F. Morrison. *Biochemistry* **1984**, *23*, 2753.

(18.) Coordinates made available from D. A. Matthews in prerelease form.

(19.) R. L. Blakely *The Biochemistry of Folic Acid and Related Pteridines*, North-Holland Publ.: Amsterdam (1969).

(20) BIOGRAF is the interactive molecular mechanics/graphics program from Molecular Simulations Inc. (BioDesign) , Pasadena, CA. All simulations were performed on DEC VAX 8650 and Alliant FX8/8 computers, and structures were displayed on Evans & Sutherland PS330 and PS390 series graphics systems.

(21.) F. M. Richards. *Ann. Rev. Biophys. Bioeng.* **1977**, *6*, 151.

(22.) M. L. Connolly. *J. Appl. Cryst.* **1983**, *16*, 548.

(23.) A. M. Naylor, W. A. Goddard III. In *Impact of Surface and Interfacial Structure on Enzyme Activity*, J. D. Burrington and D. Clark, Eds. (ACS Symposium Series: Washington, DC, 1989) in press.

(24.) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(25.) L. Cocco, C. Temple, Jr., J. A. Montgomery, R. E. London, R. L. Blakley. *Biochem. Biophys. Res. Commun.* **1981**, *100*, 413.

(26.) W. L. Jorgenson. *J. Amer. Chem. Soc.* **1981**, *103*, 335.

(27.) L. X. Dang, B. M. Pettitt. *J. Phys. Chem.* **1987**, *91*, 3349.

(28.) J. Labanowski, I. Motoc, C. B. Naylor, D. Mayer, R. A. Dammkoehler. *Quant. Struct.-Act. Relat.* **1986**, *5*, 138.

Appendix I

A variety of force-field methods have evolved in the last few years.[1-7] Their differences are based on the potential functions and parameter sets used for the calculations. Typically, the energy function is based upon the separation of internal coordinate and nonbond interaction terms. We will focus our discussion on the specific functions used in the AMBER force-field[2,3].

The total energy for the system, E_{tot}, is given by:

$$E_{tot} = E_b + E_\theta + E_\phi + E_\omega + E_{vdW} + E_{el} + E_{Hb}, \qquad (AI-1)$$

where E_b, E_θ, E_ϕ and E_ω are those contributions that are due to covalent terms and that correspond to bonds, angles, torsions, and inversions, respectively. The non-bond components of the energy are due to van der Waals, electrostatic, and hydrogen-bond interactions; these are represented in Equation (AI-1) by E_{vdW}, E_{el}, and E_{Hb}, respectively.

For many systems, it is reasonable to consider the molecule as a set of masses held together by springs. Hooke's Law may then be used to calculate the energy contribution involved with bond and angle distortions from their natural values (E_b and E_θ, respectively):

$$E_b = \frac{1}{2}\sum_{i=1}^{N} k_i^b (r_i - r_i^0)^2 \qquad (AI-2)$$

$$E_\theta = \frac{1}{2}\sum_{i<j}^{M} k_{ij}^\theta (\theta_{ij} - \theta_{ij}^0)^2, \qquad (AI-3)$$

where N and M are the total number of bonds and bond angles, respectively, in the molecule. The k^bs and k^θs are the force constants determined empirically,

the r_i and θ_{ij} are the actual bond lengths and angles, and r_i^0 and θ_{ij}^0 are the natural bond lengths and angles for the unconstrained system.

Rotational barriers must be built into the force field:

$$E_\phi = \sum_m \{|k_{m\phi}| - k_{m\phi}\cos(n\phi)\}, n = 1,2,3,4,5,6. \qquad (AI-4)$$

This Fourier series describes the change in energy associated with change in torsion angle ϕ. The force constant, $k_{m\phi}$, may be negative for a component, indicating a preference for the trans conformation of a group of substituents. The n's account for different rotational periodicities.

Umbrella forces (inversions), E_ω, are incorporated to describe the nonplanar or planar arrangement of trigonally bonded atoms:

$$E_\omega = \sum_{i=1}^{O} k_\omega(\omega_i - \omega_i^0)^2. \qquad (AI-5)$$

This supplement to the torsional potential provides a better description near the minimum energy geometry and is important for molecular dynamics analysis.

The potential energy function used to describe the attractions of London dispersion forces and the repulsions as the atoms get too close together is the Lennard-Jones 12-6 potential:

$$E_{vdW} = \sum_{i<j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}\right)\cdot SW. \qquad (AI-6)$$

Here, the A_{ij} and B_{ij} are constants derived from the position and depth of the minimum of the interaction potential, r_{ij} is the distance between atoms i and j, and SW is a switching function used to smoothly truncate nonbonded contributions.

A Coulomb potential is used to account for the electrostatic interactions:

$$E_{el} = \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\cdot SW, \qquad (AI-7)$$

where q_i represents the charge on atomic center i, ϵ_i is the dielectric constant, and r_{ij} is the distance between atoms i and j. Once again, a switching function is employed to truncate contributions smoothly to the electrostatic energy of the system.

The contribution to the energy from hydrogen bonds, E_{Hb}, is determined via a 12-10 potential function of the form:

$$E_{Hb} = \sum_{H-bonds} \left(\frac{C_{ij}}{r_{HA}^{12}} - \frac{D_{ij}}{r_{HA}^{10}}\right)\cos^m(\theta_{A-H-D}) \cdot \cos^n(\theta_{AA-A-H}) \cdot SW' \cdot SW''. \qquad (AI-8)$$

The parameters C_{ij} and D_{ij} are determined, based upon the minimum of the interaction potential, and r_{HA} is the distance between the hydrogen and its H-bond acceptor heteroatom. θ_{A-H-D} represents the angle between the H-bond acceptor, hydrogen, and donor atoms, while θ_{AA-A-H} describes that angle defined by the atom preceding the acceptor, the acceptor, and hydrogen atoms. As above, switching functions are used for smooth truncation.

References

(1) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus. *J. Comp. Chem.*, **1983**, 4, 187.

(2) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner. *J. Am. Chem. Soc.*, **1984**, 106, 765.

(3) S. J. Weiner, P. A. Kollman, D. T. Nguyen, P. A. Case. *J. Comp. Chem.*, **1986**, 7, 230.

(4) U. Burkert, N. L. Allinger. *Molecular Mechanics*, ACS Monograph 177, **1982**, Washington, D. C.

(5) A. Hagler, E. Euler, S. Lifson. *J. Amer. Chem. Soc.*, **1974**, 96, 5319.

(6) F. Momany, R. McGuire, A. Burgess, H. Scheraga. *J. Phys. Chem.*, **1975**, 79, 2361.

(7) BIOGRAF Reference Manual, BioDesign, Inc. Pasadena, CA 91101.

Appendix II

The text of this appendix is a communication coauthored with William A. Goddard III, Garry E. Kiefer, and Donald A. Tomalia and has been submitted to the *Journal of the American Chemical Society.*

Starburst Dendrimers V: Molecular Shape Control

Adel M. Naylor and William A. Goddard III*

Contribution No. 7786

from the Arthur Amos Noyes Laboratory of Chemical Physics,

California Institute of Technology, Pasadena, California 91125

and

Garry E. Kiefer and Donald A. Tomalia*

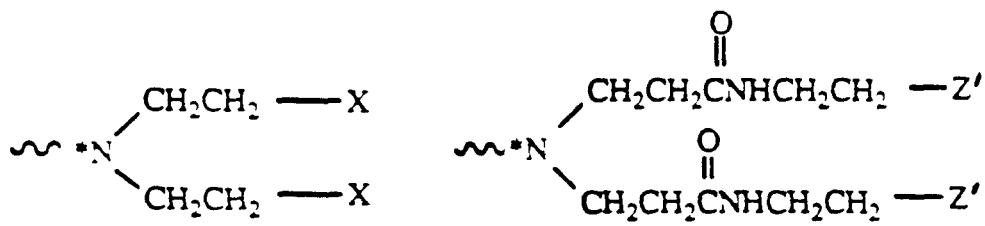
Functional Polymers Research

Dow Chemical U.S.A., Midland, Michigan 48640

Molecular simulation studies show that starburst dendrimers may be shape-designed to produce a continuum of morphologies ranging from open hemispherical domes to closed spheroids possessing internal hollows and channels. Carbon-13 spin-lattice relaxation times (T_1) of various guest molecules parallel these morphological developments, supporting the notion of unimolecular encapsulation.

The recently reported "starburst dendrimer"[1,2] and "arborol"[3] systems are unimolecular prototypes that allow for both *endo*-receptor and *exo-supramolecular* controlled molecular morphogenesis. This present communication describes the use of molecular simulation and reiterative generation development as a means for controlling molecular size and topology.

In a previous communication[2] where the initiator core (**I**) = NH_3 and the interior branch segments are β-alanine units such as (**2**),



(1) **(2)**

$Z' = NH_2$, $X = CO_2Me$, * = branch juncture

it was presumed that these symmetrical junctured dendrimers were expanding three-dimensionally to produce regular spheroids as a function of generation (see Figure 1). Detailed experimental structural data were not available on these materials due to the fractal nature of the polymers. Hence, in order to provide information on the overall shape of these dendrimers and on the presence and character of internal guest/host sites, we carried out molecular mechanics simulations of generations 1 through 7.[4]

The calculations used the AMBER force field[5] with the **POLYGRAF** molecular simulation programs.[6] These simulations indicate a dramatic change in morphology with generation, as indicated in Table I. Generations 1-3 are highly asymmetric (I_z/I_x = 4.4 to 2.7), whereas generations 5-7 are nearly spherical (I_z/I_x = 1.3), with the transition between the two forms (I_z/I_x = 1.7) occur-

ring at generation 4. We find the average structures for the early generations to be very open, domed shapes, as shown in Figure 2. The latter generations are more dense, spheroid-like topologies (Figure 2c) with solvent-filled interior hollows connected by channels that run the entire length of the macromolecule (Figure 2d). Generations 4 through 7 seem quite capable of encapsulating host molecules.

To experimentally investigate both the morphological changes as a function of generation and the predicted internal voids in the higher generation dendrimers, we carried out NMR studies of guest molecules in solution with carbomethoxy-terminated β-alanine dendrimers. Using the method of Menger and Jerkunica,[7] we found that 2,4-dichlorophenoxy acetic acid **(3)** or acetylsalicylic acid **(4)** exhibit behavior in the presence of dendrimer host molecules that supports the morphological changes predicted by molecular simulation. For example, adding ~12-68 mmol of **(3)** or **(4)** to 1 mmol of the starburst carbomethoxy-terminated dendrimers (A); (generations 0.5-5.5) in $CDCl_3$ produced spin lattice relaxation times (T_1's) for designated carbons that decrease from a maximum at generation 0.5 to a minimum at generation 3.5 (see Table II and Figure 3). The stoichiometries for guest:host were shown to be ~4:1 by weight and ~3:1 based on a molar comparison of dendrimer interior tertiary nitrogen:guest carboxylic acid moieties for generations = 2.5-5.5. Exceeding these guest:host ratios produces resonance signals with T_1 values characteristic of the guest molecules in bulk solvent phase in the absence of the dendrimeric hosts.

The T_1's for the small organic molecules reach a minimum at dendrimer generation 3.5, while the simulations show a transition into a spheroid-like morphology between generations 3 and 4. This is indicated in Figure 3, where the maximum asymmetry (I_z/I_x) is plotted as a function of generation. We suggest

that this occurs because these macromolecules (generation 4 and above) are able to effectively encapsulate the smaller guest molecules and hence decrease the measured T_1's. This minimization of T_1's does not necessarily mean that the guest molecules are totally encapsulated; rather, they could be aggregated and congested at the surface of the polymer matrix. Thus we conclude that the suppressed T_1 values (G = 3.5-5.5) are supportive of the extensive channeling and presence of interior void space predicted by the molecular dynamics simulations.

In conclusion, this work describes molecular simulations and physical measurements that suggest starburst branching strategies may be useful for controlling molecular morphogenesis.

ACKNOWLEDGMENTS:

We wish to gratefully acknowledge useful discussions with Drs. D. M. Hedstrand and L. R. Wilson. The simulations were funded by a contract from Energy Conversion and Utilization Technologies Program of the Department of Enerrgy. The computer equipment used was funded by a contract (No. N00014-86-K-0735) from Defense Advanced Research Projects Agency, Office of Naval Research, and by a grant (No. DMR-84-21119) from the Division of Materials Research of the National Science Foundation, Materials Research Groups.

REFERENCES AND NOTES

(1) Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, M.; Kallos, G.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. *Polym. J. (Tokyo)* **1985,** *17,* 117-132; *Macromolecules* **1986,** *19,* 2466; Tomalia, D. A.; Berry, V.; Hall, M.; Hedstrand, D. M. *Macromolecules* **1987,** *20,* 1164; Hall, H.; Padias, A.; McConnell, R.; Tomalia, D. A. *J. Org. Chem.* **1987,** *52,* 5305; Tomalia, D. A.; Dewald, J. R. U. S. Patent 4 507 466, 1985; U. S. Patent 4 558 120, 1985; U. S. Patent 4 568 737, 1986; U. S. Patent 4 587 329, 1986; U. S. Patent 4 631 337, 1986; U. S. Patent 4 694 064, 1987.

(2) Tomalia, D. A.; Hall, M.; Hedstrand, D. M. *J. Am. Chem. Soc.* **1987,** *109,* 1601.

(3) Newkome, G. R.; Yao, Z.-q.; Baker, G. R.; Gupta, V. K. *J. Org. Chem.* **1985,** *50,* 2003; Newkome, G. R.; Yao, Z.-q.; Baker, G. R.; Gupta, V. K.; Russo, P. S.; Saunders, M. J. *J. Am. Chem. Soc.* **1986,** *103,* 849; Newkome, G. R.; Baker, G. R.; Saunders, J. J.; Russo, P. S.; Gupta, V. K.; Yao, Z.-q.; Miller, J. E.; Bouillion, K. *J. Chem. Soc., Chem. Commun.* **1986,** 752.

(4) Naylor, A. M.; Goddard III, W. A., manuscript in preparation.

(5) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U.C.; Ghio, C.; Alagona, G.; Proteta, Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984,** *106,* 765.

(6) BioDesign, Inc., Pasadena, California, 91101.

(7) Menger, F. M.; Jerkunica, J. M. *J. Am. Chem. Soc.* **1978,** *100,* 688.

(8) Richards, F. *Ann. Rev. Biophys. Bioeng.* **1977,** *6,* 151.

Table I. Moments of inertia for β-alanine starburst dendrimers from molecular dynamics simulation (x, y, and z are the principal axes).

Generation	I_{TOT}[a]	% I_x	% I_y	% I_z	I_z/I_x
1	1.056	11.26	38.58	50.16	4.5
2	9.914	13.85	28.61	57.54	4.2
3	43.24	18.77	29.91	51.31	2.7
4	129.3	24.59	34.85	41.54	1.7
5	379.2	29.06	34.54	36.40	1.3
6	991.6	29.52	33.54	36.94	1.3
7	3131.0	29.88	31.16	38.96	1.3

[a]Units: Daltons $\times$ Å^2.

Table II. Carbon-13 spin lattice relaxation (T_1, sec.) measurements for acetylsalicyclic acid **(3)** and 2,4-dichlorophenoxy acetic acid **(4)** in the presence of dendrimers (A) where terminal groups are designated as in **(1)** with X = $-CO_2Me$.

	Designated Carbons					
Dendrimer	Guest Molecule **(3)**			Guest Molecule **(4)**		
Generation	3	5	6	3	5	6
—	1.58 ± 0.04 (132.27)[a]	1.30 ± 0.03 (134.54)	1.76 ± 0.06 (123.81)	1.66 ± 0.10 (130.41)	2.70 ± 0.45 (127.66)	1.6 (114.15)
0.5	0.82 ± 0.01 (131.67)	0.60 ± 0.01 (133.08)	0.81 ± 0.01 (123.23)	0.74 ± 0.02 (129.67)	0.89 ± 0.01 (127.28)	0.74 ± 0.02 (114.48)
1.5	0.34 ± 0.01 (131.43)	0.30 ± 0.01 (132.41)	0.37 ± 0.02 (122.87)	0.47 ± 0.004 (129.70)	0.57 ± 0.01 (127.38)	0.049 ± 0.02 (114.58)
2.5	0.26 ± 0.01 (131.63)	0.24 ± 0.01 (132.52)	0.27 ± 0.01 (123.07)	0.38 ± 0.004 (129.79)	0.43 ± 0.01 (127.54)	0.35 ± 0.02 (114.78)
3.5	0.22 ± 0.02 (131.26)	0.22 ± 0.02 (132.13)	0.22 ± 0.01 (122.76)	0.34 ± 0.01 (128.64)	0.38 ± 0.01 (127.41)	0.35 ± 0.01 (114.66)
4.5	0.23 ± 0.01 (131.42)	0.25 ± 0.01 (132.28)	0.27 ± 0.02 (122.92)	0.40 ± 0.01 (129.57)	0.39 ± 0.01 (127.35)	0.32 ± 0.02 (114.63)
5.5	0.24 ± 0.01 (131.48)	0.24 ± 0.01 (132.51)	0.26 ± 0.01 (123.00)	0.34 ± 0.01 (129.38)	0.35 ± 0.01 (127.16)	0.31 ± 0.01 (114.39)

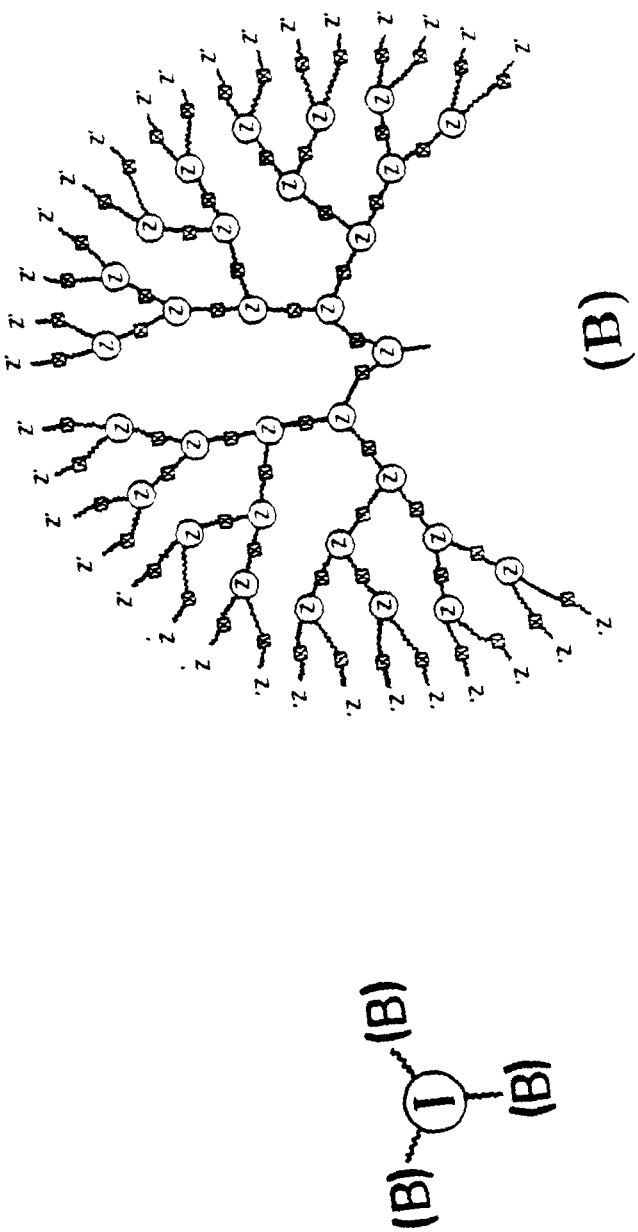
[a]Carbon-13 chemical shifts (ppm).

Figure Captions

Figure 1. (A) Tri-dendron (tridirectional) dendrimer derived from ammonia where (I) = N and (B) are symmetrical branch junctured dendrons. (B) Mono-dendron (unidirectional) starburst branching patterns where X = amide linkages are derived from **(1)**, Z = branch junctures as designated in **(1)** and **(2)** advanced to generation = 6.

Figure 2. Orthogonal views of a β-alanine dendrimer. (a) and (b) illustrate the open, domed nature of the early generations (generation 3 shown), while (c) and (d) show the spherical nature of the late generations (generation 6 shown). In (d) the character of the solvent-filled void spaces is depicted with the use of solvent-accessible surfaces.[8].

Figure 3. (Left Axis) Comparison of carbon-13 spin lattice relaxation times (T_1, sec.) versus generation for designated carbons (3, 5, and 6; ▲, □, and ○, respectively) in **(3)** with $CHCl_3$ solvent and various dendrimer generations. (Right Axis) Molecular asymmetry (- - - ● - - -) versus dendrimer generation (number of Z′ groups).



(A) (B)

Figure 1.

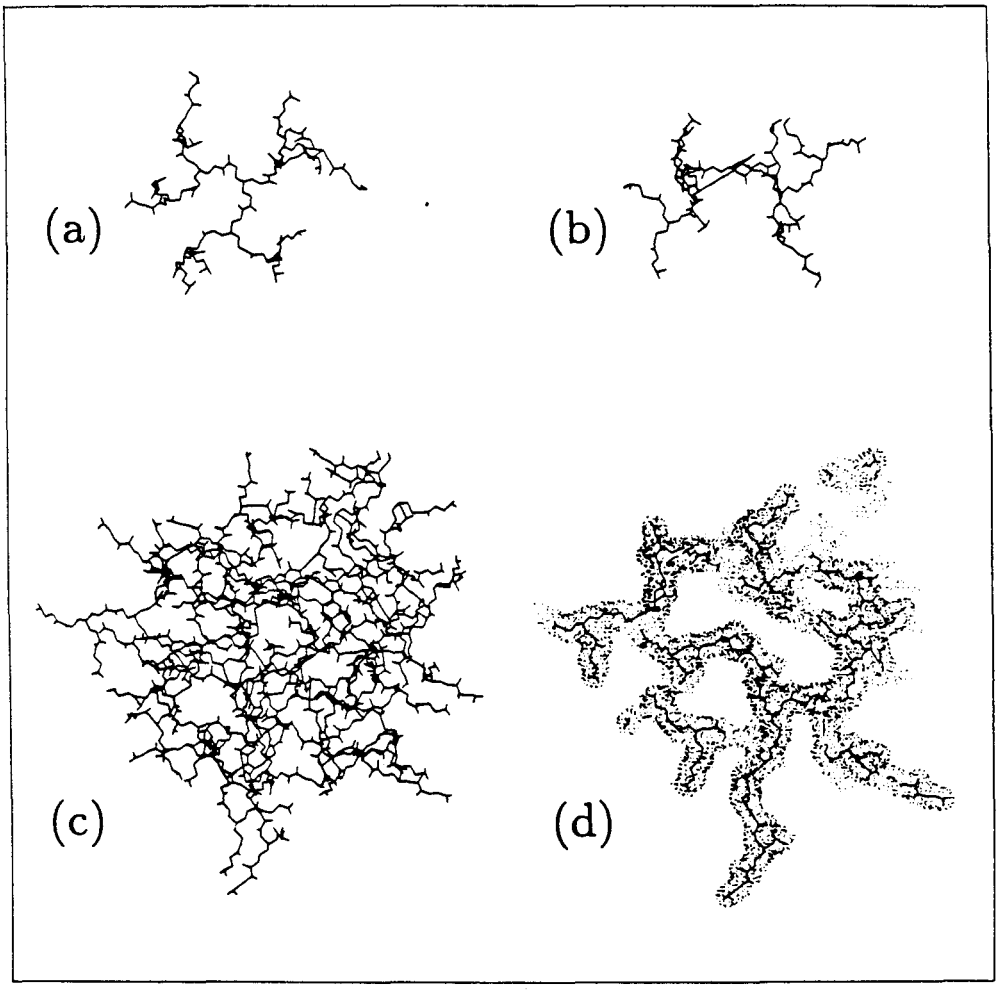


Figure 2.

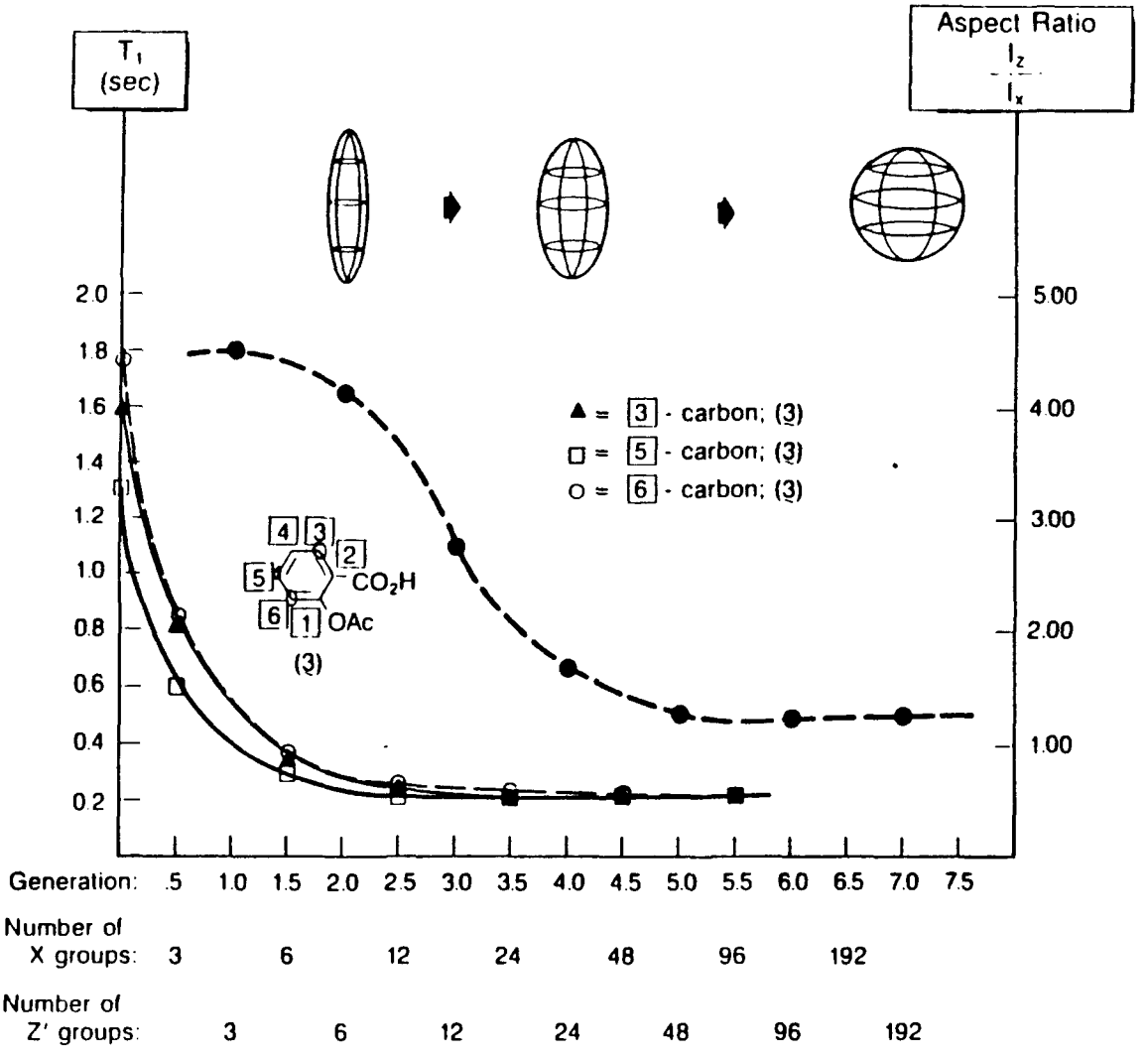

T1 (sec)
Aspect Ratio
Iz / Ix
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
5.00
4.00
3.00
2.00
1.00
▲ = 3 - carbon; (3)
□ = 5 - carbon; (3)
○ = 6 - carbon; (3)
CO2H
OAc
(3)
Generation: .5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5
Number of X groups: 3 6 12 24 48 96 192
Number of Z' groups: 3 6 12 24 48 96 192

Figure 3.

Appendix III

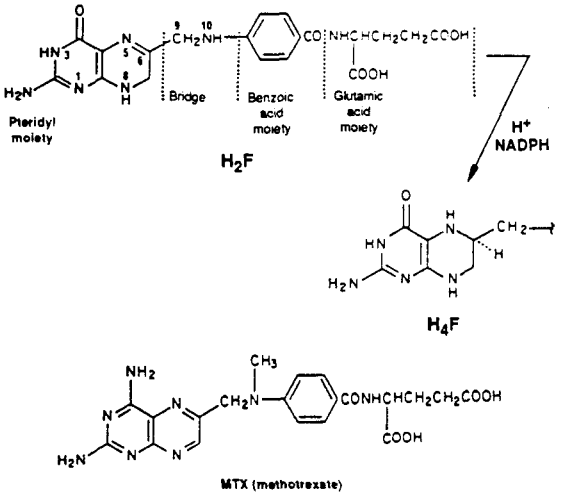
The text of this appendix is an article coauthored with Stephen J. Benkovic and Carol A. Fierke, which appeared in *Science*.

Insights into Enzyme Function from Studies on Mutants of Dihydrofolate Reductase

S. J. BENKOVIC, C. A. FIERKE, A. M. NAYLOR

Kinetic analysis and protein mutagenesis allow the importance of individual amino acids in ligand binding and catalysis to be assessed. A kinetic analysis has shown that the reaction catalyzed by dihydrofolate reductase is optimized with respect to product flux, which in turn is predetermined by the active-site hydrophobic surface. Protein mutagenesis has revealed that specific hydrophobic residues contribute 2 to 5 kilocalories per mole to ligand binding and catalysis. The extent to which perturbations within this active-site ensemble may affect catalysis is discussed in terms of the constraints imposed by the energy surface for the reaction.

DIHYDROFOLATE REDUCTASE (5,6,7,8-TETRAHYDROFOLATE: NADP oxidoreductase, E.C. 1.5.1.3) catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate (H_2F) to 5,6,7,8-tetrahydrofolate (H_4F). The enzyme is necessary for maintaining intracellular pools of H_4F and its derivatives, which are essential cofactors in the biosynthesis of purines, thymidylate, and several amino acids. It is the target enzyme of a group of antifolate drugs, methotrexate (MTX), trimethoprim, and pyrimethamine, that are widely used as antitumor and antimicrobial agents (*1*).



The identities of the amino acids at the active site of dihydrofolate reductase (DHFR) are now known, but the function and relative importance of the amino acid residues in binding and catalysis have not been clarified. We have used site-directed mutagenesis to establish structure-function relations for the *Escherichia coli* DHFR and in particular to evaluate the importance of hydrophobic amino acids on the function of this enzyme and on the binding of inhibitors.

The nature of the question posed presupposes two important sets of data (*2*): (i) an accurate structure that closely approximates the enzyme·H_2F·NADPH reactive ternary complex so that the key amino acid contacts are revealed and (ii) the elucidation of a complete kinetic pathway that can explain the steady-state behavior of DHFR in terms of the rate-limiting steps in turnover under a variety of conditions and that allows for direct measurement of the catalytic step at the enzyme-active site.

Structure

Visualization of the active-site amino acids was provided by the x-ray diffraction studies of Matthews *et al.* (*3*) on crystals of the *E. coli* DHFR·MTX complex and of the *Lactobacillus casei* DHFR·MTX·NADPH ternary complex (*4*, *5*) refined to 1.7 Å (*6*, *7*). The two backbone structures are similar, despite less than 30% homology for the two sequences; when the coordinates of 142 of the α-carbon atoms (out of 159) of the *E. coli* DHFR are matched to the structurally equivalent carbons of the *L. casei* DHFR, the root-mean-square deviation is only 1.07 Å (*6*). The active site of both proteins is a cavity ~15 Å deep that is lined by hydrophobic side chains; the only polar side chain is the carboxyl group of Asp^{27} (*E. coli*) (Fig. 1). NADPH binds in an extended conformation with the nicotinamide moiety inserted through the entrance of the cavity where it makes hydrophobic and hydrogen-bonding interactions with residues on the bottom and sides of the site. Methotrexate binds in an open conformation with the pteridine ring nearly perpendicular to the benzoyl ring. The pyrimidine edge of the pteridine ring is deeply buried in the active site, whereas the glutamate side chain extends out of the cavity to the surface. Selected key hydrophobic contacts are also illustrated in Fig. 1. The same interactions have been built into the proposed enzyme·NADPH·H_2F complex (*6*) but with the pteridine ring turned over with respect to the side chain by a 180° rotation about the C6–C9 bond, as demanded by the stereochemical course of the reduction (*8*). However, Naylor and Goddard (*9*) have suggested, with the use of molecular mechanics calculations, that the energetics of these alternate forms are nearly equal (within a few kilocalories) for both H_2F and MTX bound to DHFR. The two complexes serve as the basis for our selection of residues for mutation.

S. J. Benkovic is in the Department of Chemistry, Pennsylvania State University, University Park, PA 16802. C. A. Fierke is in the Department of Biochemistry, Duke University Medical Center, Durham, NC 27710. A. M. Naylor is in the Department of Chemistry, California Institute of Technology, Pasadena, CA 91125.

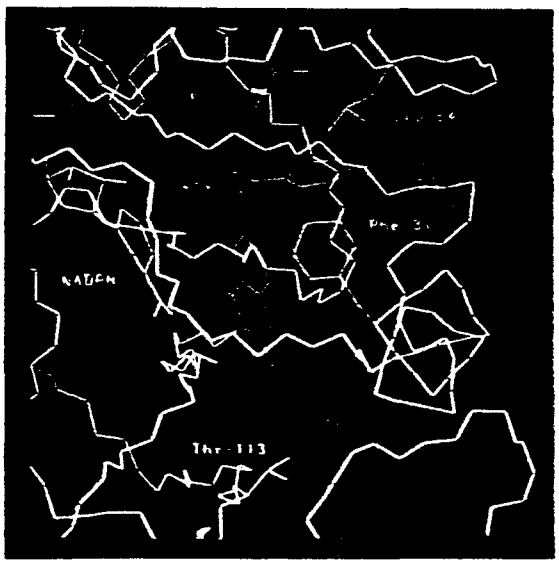


Fig. 1. Carbon backbone structure of *E. coli* DHFR including bound ligands (MTX and NADPH) and the side chains for key amino acids [depicted with BIOGRAF (*25*)].

Kinetics

The kinetic sequence for *E. coli* DHFR that predicts steady-state kinetic parameters and full-time course kinetics under a variety of substrate concentrations and *p*Hs was established by measuring the various ligand association and dissociation rate constants and pre-steady-state reaction transients with the use of stopped-flow fluorescence and absorbance spectroscopy (*10*). The scheme at 25°C is shown in Fig. 2, where NH = NADPH and N = NADP. Several key features relevant is our discussion are: (i) steady-state turnover is limited at *p*H 6 to 7 by H_4F release; (ii) dissociation of H_4F is fastest from the mixed ternary E·NH·H_4F complex; consequently the kinetic pathway for steady-state turnover at saturating substrate concentrations follows a specific, preferred pathway, in which H_4F dissociation from DHFR occurs after NADPH replaces NADP so that the catalytic cycles are interlocked (heavy arrows); and (iii) the reaction strongly favors H_4F formation (overall equilibrium constant $K_{ov} \approx 10^4$, *p*H 7) that is partially reflected in the internal equilibrium E·NH·H_2F ⇌ E·N·H_4F (internal equilibrium constant $K_{int} \approx 10^3$, *p*H 7) for the reactive ternary complexes.

Examination of the *p*H dependence of these reactions established an active-site group with an intrinsic pK_a of 6.5 for the E·NH·H_2F ternary complex—a value that is the same for all forms of the enzyme: free, binary, and ternary complexes (*11–13*). Protonation of this group is required for the chemical step that involves hydride transfer from NH to H_2F. This active-site group is most likely Asp^{27}. Replacement of the Asp^{27} by Asn or Ser generates new enzyme forms that do not have an acidic proton at their catalytic site and that require preprotonated H_2F for activity (*14*). However, because of the location of the carboxyl of Asp^{27}, its proton cannot directly transfer to N5 of H_2F and requires a pathway through intervening water molecules or possibly O4 of H_2F (*15*). Although the proton transfer component of the reduction step is thermodynamically unfavorable [pK_a (Asp^{27}) = 6.5; pK_a (N5 of H_2F) = 3.8 (*16*)], the calculated proton transfer rate, 2×10^7 sec^{-1}, provides a sufficient concentration of protonated H_2F so that the observed rate for reduction (950 sec^{-1}) is not rate-limiting.

Free Energy Profiles

The detailed descriptions of reactions in terms of free energy profiles have provided impressive insights into the coupling between the overall thermodynamic free energy change for the given reaction and specific steps in the kinetic sequence for enzyme turnover (*17, 18*). The free energy profiles provide a quantitative comparison between the ground and transition state levels for enzyme-ligand complexes (*19–21*).

Before we consider the effect of site-specific mutations on the characteristics of DHFR, we evaluate the efficiency of this enzymic process. Given the definition of the kinetic sequence under physiological conditions (we introduce concentrations for NH, N, H_4F, and H_2F that approximate those of the *E. coli* cell), we can calculate the steady-state turnover under these conditions. A hypothetical maximum turnover is set by having as the rate-limiting step the diffusional binding of H_2F to E·NH (*22*), because the higher concentration of NH would maintain the reductase in the E·NH form. Under these conditions the enzyme is operating at 15% of its hypothetical maximum turnover, slightly limited by the dissociation of H_4F from E·NH·H_4F. [This efficiency could be less if, under cellular conditions, channeling of H_2F to the enzyme occurred (*23*).] Consequently, the division of the free energy change for the DHFR-catalyzed reaction into the steps in Fig. 3 is a satisfactory solution to the problem of catalyzing this process intracellularly.

There are a number of pathways that efficiently solve the energetics of the reduction within the thermodynamic constraints imposed by the overall reaction equilibrium and the diffusional barrier of substrate-enzyme combination. All of these solutions have two features in common: (i) the energy levels of the reaction intermediates are above that of the product ground state and (ii) the intermediate transition state energy levels are below those for forming the E·NH·H_2F complex.

One solution is equivalent to stabilizing equally all of the intermediate ground and transition states relative to a primitive reaction pathway, a process designated as uniform binding (*17*). The DHFR kinetic scheme conforms to within 1.6 kcal mol^{-1} to the predictions of the uniform binding constraint. Moreover, the internal states, E·NH·H_2F and E·N·H_4F, need not be at the same energy level for high catalytic efficiency (*21, 24*) because of the apparent irreversibility of the overall conversion. Simulations predict that this internal equilibrium constant can vary by 10^3 with less than a twofold change in turnover efficiency when the uniform binding constraint is maintained (*20*). An important consequence of the high efficiency of the wild-type enzyme is that mutations generally will decrease the flux of the reductase reaction under physiological conditions.

Comparison of *E. coli* and *L. casei* DHFRs

We have studied two kinds of mutants, those that occur naturally as in the comparison of the *E. coli* and *L. casei* enzymes and those generated through site-specific mutagenesis at the active site of the *E. coli* enzyme.

The active-site amino acids of the two bacterial DHFRs were overlaid with MTX as a center for alignment; their solvent-accessible surfaces were displayed with BIOGRAF (Fig. 4) (*25*). All of the residues that contained any atom within 5 Å of MTX were included. The active-site amino acid homology was computed to be 55% from a base ensemble of 40 amino acids; those that were either identical or have only backbone interactions were deemed homologous. As a calibration point, the solvent-accessible surface of MTX in the binary *E. coli* and ternary *L. casei* complexes is identical (±2%).

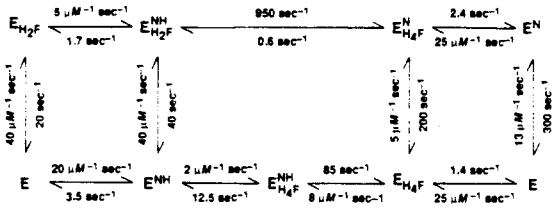


Fig. 2. Kinetic scheme for dihydrofolate reductase at 25°C, MTEN buffer; N, NADP; NH, NADPH; H_2F, dihydrofolate; H_4F, tetrahydrofolate.

Comparison of the total area of the solvent-accessible surfaces of the two proteins within 5 Å of the MTX revealed that the *E. coli* area is 93% that of *L. casei* (after deletion of the residues 15 to 20 in *L. casei* that are positioned by NADP). The congruence in the protein surface in the region of MTX is remarkably similar; discernible differences occur, however, near the glutamate and benzoyl residues (Fig. 4).

Consistent with the apparent similarities in protein-substrate interactions are the kinetic schemes of the two enzymes, which are nearly identical and feature rapid, essentially irreversible hydride transfer followed by rate-limiting H_4F dissociation from E·NH·H_4F, as described previously for the *E. coli* enzyme. A comparison of the reaction coordinate diagrams for the two enzymes under physiological conditions is shown in Fig. 3 (*26*). The major difference observed is that the free energy for binding NADPH to the *L. casei* enzyme is 2 kcal mol^{-1} more favorable than the *E. coli* enzyme, whereas the remainder of the free energy differences are less than 1 kcal mol^{-1}. We presume that NADPH is positioned identically on both enzymes.

We conclude that the energy levels of ground and transition state species are predetermined by interactions between the substrate and the active-site surface, which can be constructed by a variable combination of amino acids. This analysis does not preclude specific transition state stabilization (*27, 28*) that arises from changes on the protein surface as the reaction proceeds; obviously both the DHFR from *E. coli* and *L. casei* can respond similarly. The design of species-specific inhibitors should exploit specific regions where obvious differences are manifest, that is Ile^{50} (*E. coli*) versus Phe^{49} (*L. casei*).

Site-Specific Mutants

In the study of designed mutant proteins, an important initial question to address is their structural integrity. Recent crystallographic studies indicate that point mutations are in general accommodated by very minor readjustments of the tertiary protein structure with water molecules occupying the space created by a smaller amino acid α side chain (*14, 29, 30*). In the case of the *E. coli* DHFR, x-ray crystallographic studies of the Asn^{27} and Ser^{27} mutants showed that neither the MTX binding geometry nor the detailed three-dimensional topography of the enzyme was altered by the mutation (*14*). A kinetic analysis of the conformational states of various reductase mutants generated at the sites depicted in Fig. 1 established that the distribution and rate of interconversion of the two conformational states of the free and mutant enzymes were generally unchanged (*31–33*). We conclude that our mutants have the same basic conformation as the wild type and that changes in their binding and catalytic properties can be attributed primarily to the altered amino acid, but we note that substitution leads to localized conformational changes as the protein structure readjusts (*34*) and thus requires a cautious interpretation of any data (vide infra).

We consider mutations at three strictly conserved amino acids (Phe^{31}, Leu^{54}, and Thr^{113}). Phe^{31} resides in a hydrophobic pocket and interacts with the pteroyl moiety of H_2F through van der Waals contacts such that the edge of the phenyl ring is oriented toward faces of both the pteridine ring and *p*-aminobenzoyl group (*7*) in an edge-to-face aromatic-aromatic interaction (*35*). The isobutyl side chain of Leu^{54} is part of a hydrophobic region around the benzoyl glutamate portion of the H_2F but is more than 10 Å distant from the site of hydride transfer. The hydroxyl of Thr^{113} forms a hydrogen bond to the carboxylate oxygen of Asp^{27} and interacts with the 2-amino group of the pteridine moiety through a hydrogen-bonded water, 201. The effects of these mutations on key rate steps in dihydrofolate reductase turnover are exhibited in Table 1. There are a number of salient changes: (i) the dissociation constants for H_2F and H_4F increase with a diminution in the bulk of the α substituent at Phe^{31} and Leu^{54} as well as the loss of the hydrogen-bonding interaction at Thr^{113}; (ii) the maximum velocity V_M increases by a factor of 2 to 3 as a consequence of increasing k_{off} (the dissociation rate constant of H_4F from E·NH·H_4F); (iii) the rate constant for the hydride transfer step (k_H) decreases with the Gly^{54} and the [Gly^{54}, Val^{31}] mutants so that it is rate-limiting for turnover; and (iv) increases in the dissociation constants for MTX qualitatively parallel those for H_2F and H_4F. Although these mutant enzymes have larger V_M values, the smaller binding constants for H_2F ($1/K_{H_2F}$) lower their turnover in a cellular milieu where the concentration of H_2F is estimated to be 0.3 μ*M*.

From a comparison of various mutants to the wild-type protein one may estimate the contribution of a side chain group from the free energy of substrate or inhibitor binding (*36*). For example, the replacement of Leu^{54} with Gly^{54}, which is the closest approach to the complete deletion of the isobutyl side chain, may provide an estimate of the hydrophobic binding energy of this side chain. A set of these free energy changes is shown in Table 2. A difficulty with this approach is that the deletion will most certainly alter the architecture within the active site. For example, the effect of a glycine substitution is to increase the degrees of freedom of neighboring residues, and in addition, water molecules may now begin to occupy the cavity. More rigorously, let $G_E^{\ddagger}$ and $G_C^{\ddagger}$ be the free energy of the enzyme and the enzyme-substrate complex, respectively, for the equilibrium binding of A to wild-type (N = Y) and mutant (N = H) DHFRs so that:

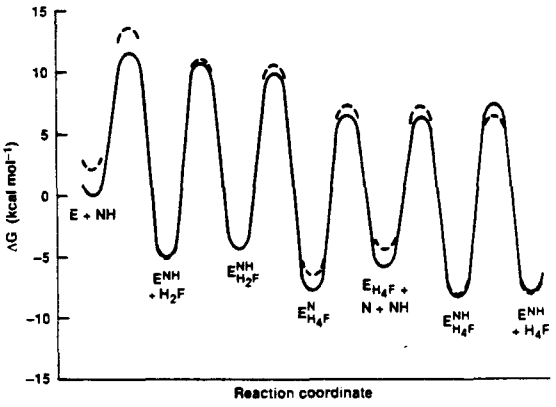


Fig. 3. Gibbs free energy coordinate diagrams for *E. coli* (———) and *L. casei* (– – –) dihydrofolate reductase aligned at the substrate ternary complex, E·NH·H_4F, calculated for conditions approximating a bacterial cell: 1.0 m*M* NADPH, 1.5 m*M* NADP, 0.3 μ*M* H_2F, 13 μ*M* H_4F, and 0.1 *M* NaCl, *p*H 7.0, 25°C.

$$E^Y + A \rightleftharpoons E^Y \cdot A \qquad \Delta G_1 = G_C^Y - G_E^Y - G_A \quad (1)$$

$$E^H + A \rightleftharpoons E^H \cdot A \qquad \Delta G_2 = G_C^H - G_E^H - G_A \quad (2)$$

$$\text{Then: } \Delta(\Delta G)_{obs} = G_C^Y - G_E^H - G_E^Y + G_E^H = \Delta G_C^{Y,H} - \Delta G_E^{Y,H} \quad (3)$$

The ground-state free energies of the enzymes may or may not be identical ($\Delta G_E^{Y,H} \neq 0$), nor is it obvious when this term can be safely neglected; thus $\Delta\Delta G_{obs}$ may not equal the intrinsic binding energy $\Delta G_C^{Y,H}$ contributed by group Y. A similar situation prevails for the measurement of hydrogen bond strength, where one must distinguish between apparent and intrinsic binding energies (*37*).

In terms of apparent binding energies, Leu^{54} contributes 3.9 and 5.0 kcal mol^{-1} to the respective binding of H_2F and MTX to the *E. coli* DHFR, emphasizing the greater importance of hydrophobic interactions in general (Table 2) (*38*) relative to salt bridge formation between the 2-amino group of MTX and Asp^{27}, which has been estimated to be ~1.8 kcal mol^{-1} (*14*). The same ordering, that is, hydrophobic forces are greater than electrostatic forces, for MTX binding emerges from free energy perturbation calculations (*39*). The free energy changes for the Phe^{31} and Val^{31} mutant additionally include loss of the edge-to-face weakly polar interaction between the aromatic side chain and the ligand, and in fact may be the major factor contributing to this $\Delta\Delta G$ of 2 to 3 kcal mol^{-1}. In the case of Tyr^{31}, the loss of binding affinity may reflect difficulties in solvating as well as accommodating the hydroxyl in the binary complex. The fact that $\Delta\Delta G$ for the double mutant is less than the sum of $\Delta\Delta G$ for the two single mutants indicates the absence of large structural unfolding (*40*).

Free Energy Relations

Although these changes are made at three spatially separate loci on the surface of the active site, they can be related by means of a linear free energy structure-function correlation of the type $k = CK^\beta$ that links thermodynamic with kinetic changes. In particular Fersht and his co-workers have applied these relations to mutants of tyrosyl–transfer RNA (tRNA) synthetase (*41*). In principle the free energy change between the reactant ground state and transition state ($\Delta\Delta G^+$) is correlated to the corresponding free energy change for the reactant-product equilibrium ($\Delta\Delta G$), relative to a standard reactant. Generally such plots are linear with slopes between zero and one. In the present case, however, the slope of a plot of log k_H (the rate of hydride transfer) versus log k_H/k_{-H} (the internal equilibrium) for a series of DHFR mutants is nearly vertical (>>1), suggesting that the transition state stability is sensitive to the mutation (the values of k_H decreased) but that the equilibrium involving the ground states is not. This implies that the reactant and product ground state are perturbed to the same extent—a manifestation of the uniform binding noted earlier. An alternate explanation that the perturbation is confined entirely to a particular transition state stabilization is ruled out by the effects of these mutations on the dissociation of H_2F and H_4F.

A variant on the classical free energy correlation relates changes in rate constants to changes in a second property of a common reactant. In the present case, plots of log k_H versus log K_{H_2F} (Fig. 5) are linear with slopes of unity and suggest that the perturbation caused by these mutations is twice as important in altering the free energy of the transition state for hydride transfer as compared with the ground state of the E · H_2F complex. The important point here is that the rate of hydride transfer increases in direct proportion to an increase in the affinity of the reductase for H_2F, with the wild-type enzyme affinity being maximal.

An interpretation of the collective data is that the active site is a reactive surface whose major function is to constrain the NADPH and the H_2F into an optimal reaction space and orientation. This is not a new concept; certainly the effect of extensive mutations on tyrosyl-tRNA synthetase (*42*), subtilisin (*43*), as well as alterations in substrate structure (*44*), show that all active-site groups contribute to varying extents to selective binding of substrates and to catalysis. These groups, that is, Leu^{54}, can be some distance from the reaction site and still have a pronounced influence. Moreover, the effects of the mutation should be more pronounced on the kinetic rather than the thermodynamic parameters since the former requires precise alignment of the NADPH and H_2F, whereas the latter do not. Eventually a situation can be reached when further increases in the binding affinity for H_2F do not translate into an increased rate (k_H), hence the negative deviation for the wild-type DHFR from the correlation in Fig. 5.

A key question, therefore, is the relation between the energetics of the reaction path, in this case the transfer of a hydride ion between C4 of the dihydropyridine and C6 of H_2F, as a function of the distances and angles of the three participating atoms (–C · · · H · · · C=). The construction of a molecular potential hypersurface defining the minimal energy path between the stable ground-state species either from experimental data or from theoretical calculations remains a formidable task. Chemical reaction dynamic information has been extracted from crystallographic data; two notable examples are the hydrolysis of alkylaryl acetals and the addition of nitrogen nucleophiles to carbonyls. An increase of only 0.01 Å in the ground-state length of the acetal bond to be cleaved is equivalent to a decrease of ~3 kcal mol^{-1} in the activation energy (*45*). For nucleophiles the angle of approach to the carbonyl clusters is near 105° (*46*). Houk and co-workers have located the transition states for hydride transfer from either methylamine or 1,4-dihydro-

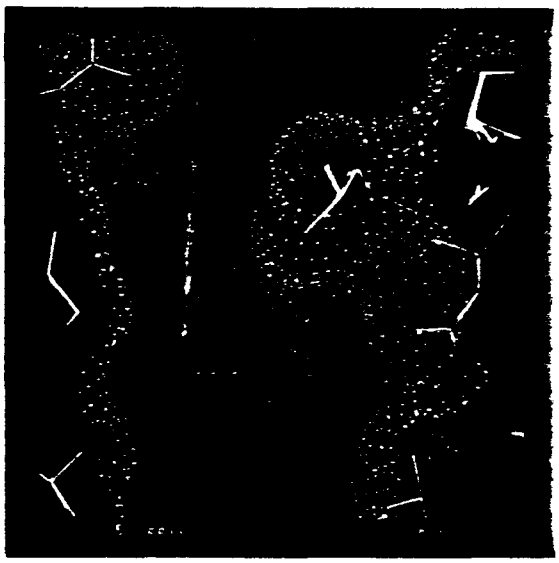


Fig. 4. Superpositioning of the active sites of the *E. coli* (gold) and *L. casei* (blue) DHFRs with MTX as a center for alignment. All residues within 5 Å of the MTX were included. Key residues that are visible in the figure (*E. coli*: *L. casei*) include Leu^{28}:Leu^{27} (upper right); Asp^{27}:Asp^{26} (middle right); Ala^7: Ala^6 (lower right); Thr^{46}:Thr^{45} (lower left); Ile^{50}:Phe^{49} (middle left); and Arg^{52}:Leu^{51} (upper left). Phe^{31} can be seen in the center of the view [depicted with BIOGRAF (*25*)].

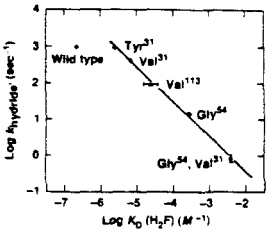


Fig. 5. Free energy correlation diagram relating the rate constant for hydride transfer to the K_D for dissociation of H_2F from a series of mutants in the folate binding site of DHFR.

Table 1. Effects of mutations on kinetic and thermodynamic parameters characteristic of *E. coli* DHFR (*10, 12, 32, 48*).

E. coli DHFR	k_H* (sec^{-1})	k_{off}* (H_4F) (sec^{-1})	V_M* (sec^{-1})	K_D (H_4F) (μM)	K_D (H_2F) (μM)	K_D (MTX) (nM)
Wild type	950	12	12	0.06	0.21	0.02
Val^{31}	400	> 20	26	0.3	6.6	3.2
Gly^{54}	14	>300	14	>15	140	100
Gly^{54}, Val^{31}	0.9	>300	0.9		2000	900
Val^{113}	120	60	32	4.0	30	0.5

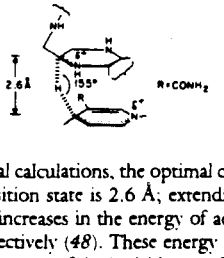
*pH-independent values in MTEN buffer, 25°C.

Table 2. Free energy changes associated with ligand binding to *E. coli* DHFR. The error for these values (SEM) is ±0.5 kcal mol^{-1}.

Mutation	Change	K_D (H_2F) $\Delta\Delta G$ (kcal mol^{-1})	K_D (MTX) $\Delta\Delta G$ (kcal mol^{-1})
Phe^{31} to Tyr	Hydrophobic	1.5	2.7
Phe^{31} to Val	Hydrophobic	2.0	3.0
Leu^{54} to Gly	Hydrophobic	3.9	5.0
Phe^{31} to Val, Leu^{54} to Gly	Hydrophobic	5.4	6.3
Thr^{113} to Val	Hydrogen bond (*2*)	2.9	1.9

pyridine to the methyliminium cation. In either case a syn transition state is favored with a bent –C · · · H · · · C= bond angle (150° to 160°) (*47*).

These conclusions may be transposed to the DHFR reaction subject to the restrictions imposed by stereochemical considerations, so that a model emerges as a possibility for the chemical reduction step.



From their theoretical calculations, the optimal carbon-carbon bond distance in this transition state is 2.6 Å; extending this distance by 0.1 or 0.3 Å causes increases in the energy of activation of 0.7 and 5.0 kcal mol^{-1}, respectively (*48*). These energy differences translate into a reduction in the rate of the hydride transfer step by factors of 3 to 5000. Thus the information gleaned from either experiment or theory argues for the importance of small molecular changes on the order of <1 Å that can account for the observed changes caused by specific mutations. On the other hand, we cannot eliminate at present the possibility that the mutated enzymes have allowed the formation of nonproductive ground-state complexes, such as bound H_2F adopting the MTX conformation (*9, 49*).

Conclusions

Insights into the importance of active-site residues in the binding of ligands and in catalysis have been obtained through the powerful combination of site-specific mutagenesis and complete kinetic analysis of the enzyme-catalyzed transformation. Hydrophobic interactions appear to dominant the binding of substrates or inhibitors as well as catalysis. The disparate mutations studied to date fit structure-reactivity correlations that linearly relate the rate of hydride transfer to the binding of H_2F, which is consistent with the hypothesis that the active-site residues act as an ensemble to create a particular surface for binding and catalyzing the reaction. The same surface can be constructed by differing combinations of amino acids, as illustrated by the *E. coli* and *L. casei* comparison, and can be unfavorably perturbed by single changes at key residues. Consequently there must be compensating changes. Information obtained from both theory and experiment document that small changes in distance may be translated into energy changes sufficient to account for our observations. These changes, on the order of <1 Å, most likely will not be detected by the present methods (x-ray and nuclear magnetic resonance) used to establish structure and these methods may only detect changes in mutated enzymes largely devoid of activity. Furthermore, one can anticipate that linearity in the structure-reactivity correlations will decay, since the correlation between the free energy of activation and reactant bond angle and distances must ultimately be nonlinear, and because the requirements for ligand binding are not as restrictive as those for catalysis.

REFERENCES AND NOTES

1. R. L. Blakley, in *Folates and Pterins* R. L. Blakely and S. J. Benkovic, Eds (Wiley, New York, 1985), vol. 2, p. 191.
2. J. R. Knowles, *Science* **236**, 1252 (1987).
3. D. A. Matthews *et al.*, *ibid.* **197**, 452 (1977).
4. D. A. Matthews *et al.*, *J. Biol. Chem.* **253**, 6946 (1978).
5. D. A. Matthews, R. A. Alden, S. T. Freer, N. Xuong, J. Kraut, *ibid.* **254**, 4144 (1979).
6. J. T. Bolin, D. J. Filman, D. A. Matthews, R. C. Hamlin, J. Kraut, *ibid.* **257**, 13650 (1982).
7. D. J. Filman, J. T. Bolin, D. A. Matthews, J. Kraut, *ibid.*, p. 13663.
8. P. A. Charlton, D. W. Young, B. Birdsall, J. Feeney, G. C. K. Roberts, *Chem. Commun.* **1979**, 922 (1979).
9. A. M. Naylor and W. A. Goddard III, unpublished results.
10. C. A. Fierke, K. A. Johnson, S. J. Benkovic, *Biochemistry* **26**, 4085 (1987).
11. S. R. Stone and J. F. Morrison, *Biochim. Biophys. Acta* **745**, 247 (1983).
12. J.-T. Chen, K. Taira, C. D. Tu, S. J. Benkovic, *Biochemistry* **26**, 4093 (1987).
13. K. Taira, J.-T. Chen, C. A. Fierke, S. J. Benkovic, *Bull. Chem. Soc. Jpn* **60**, 3025 (1987).
14. E. E. Howell, J. E. Villafranca, M. S. Warren, S. J. Oatley, J. Kraut, *Science* **231**, 1123 (1986).
15. J. E. Gready, *Biochemistry* **24**, 4761 (1985).
16. M. Poe, *J. Biol. Chem.* **252**, 3724 (1977).
17. W. J. Albery and J. R. Knowles, *Biochemistry* **15**, 5631 (1976).
18. W. P. Jencks, *Adv. Enzymol.* **43**, 219 (1975).
19. J. Stackhouse, K. P. Nambiar, J. J. Burbaum, D. M. Stauffer, S. A. Benner, *J. Am. Chem. Soc.* **107**, 2757 (1985).
20. C. A. Fierke, R. D. Kuchta, K. A. Johnson, S. J. Benkovic, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
21. J. Chin, *J. Am. Chem. Soc.* **105**, 6502 (1983).
22. M. H. Penner and C. Frieden, *J. Biol. Chem.* **262**, 15908 (1987).
23. M. Young, G. Wasserman, P. Benkovic, S. Benkovic, in *Proceedings of the Second Workshop on Folyl and Antifolyl Polyglutamates*, I. Goldman, Ed (Praeger, New York, 1985), p. 76.
24. A. D. Ellington and S. A. Benner, *J. Theor. Biol.*, in press.
25. BIOGRAF is an interactive molecular simulation three-dimensional graphics program from BIODESIGN, Inc., Pasadena, CA.
26. J. Andrews and C. A. Fierke, unpublished results.
27. L. Pauling, *Chem. Eng. News* **24**, 1375 (1946).
28. D. E. Koshland and K. E. Neet, *Annu. Rev. Biochem.* **37**, 359 (1967).
29. S. Sprang *et al.*, *Science* **237**, 905 (1987).
30. C. Chothia, *J. Mol. Biol.* **105**, 1 (1976).
31. P. J. Cayley, S. M. J. Dunn, R. W. King, *Biochemistry* **20**, 874 (1981).
32. R. J. Mayer, J.-T. Chen, K. Taira, C. A. Fierke, S. J. Benkovic, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7718 (1986).
33. J.-T. Chen, R. J. Mayer, C. A. Fierke, S. J. Benkovic, *J. Cell. Biochem.* **29**, 73 (1985).
34. J. A. Gerlt, D. W. Hibler, J. A. Wilde, P. H. Bolton, *Protein Engr.* **1**, 233 (1987).
35. S. K. Burley and G. A. Petsko, *Science* **229**, 23 (1985).

36. A. R. Fersht, in *Enzyme Structure and Mechanism* (Freeman, New York, ed 2, 1985), p. 311.
37. ———, *Trends Biochem. Sci.* **12**, 301 (1987).
38. K. Taira and S. J. Benkovic, *J. Med. Chem.*, in press.
39. C. Singh and S. J. Benkovic, unpublished results.
40. P. J. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, *Cell* **38**, 835 (1984).
41. D. M. Lowe, G. Winter, A. R. Fersht, *Biochemistry* **26**, 6038 (1987).
42. A. R. Fersht, R. J. Leatherbarrow, T. N. C. Wells, *Trends Biochem. Sci.* **11**, 321 (1986).
43. D. A. Estell *et al.*, *Science* **233**, 659 (1986).
44. C. A. Fierke and W. P. Jencks, *J. Biol. Chem.* **261**, 7603 (1986).
45. A. J. Briggs *et al.*, *J. Am. Chem. Soc.* **106**, 6200 (1984).
46. H. B. Burgi and J. D. Dunitz, *Acc. Chem. Res.* **16**, 153 (1983).
47. Y. D. Wu and K. N. Houk, *J. Am. Chem. Soc.* **109**, 2226 (1987).
48. ——— personal communication.
49. K. Taira, C. A. Fierke, J.-T. Chen, K. A. Johnson, S. J. Benkovic, *Trends Biochem. Sci.* **12**, 275 (1987).
50. We thank J.-T. Chen, R. Maver, K. Taira, K. Johnson, J. Andrews, C. Singh, K. Houk, and Y.-D. Wu for their help and involvement in this work. We also thank W. A. Goddard III for use of the computer graphics facility for graphical displays. This work was supported in part by NIH grant GM24129 and DOE-ECUT 49 242-E0403-0-3550. C.A.F. is a NIH postdoctoral fellow.

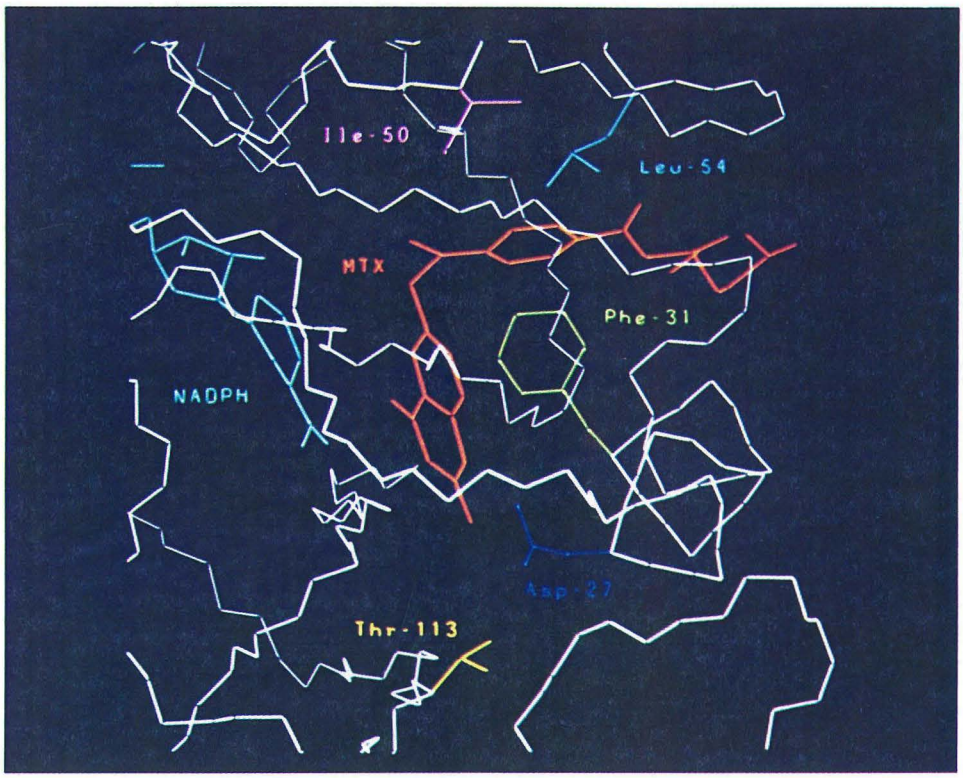


Figure 1.

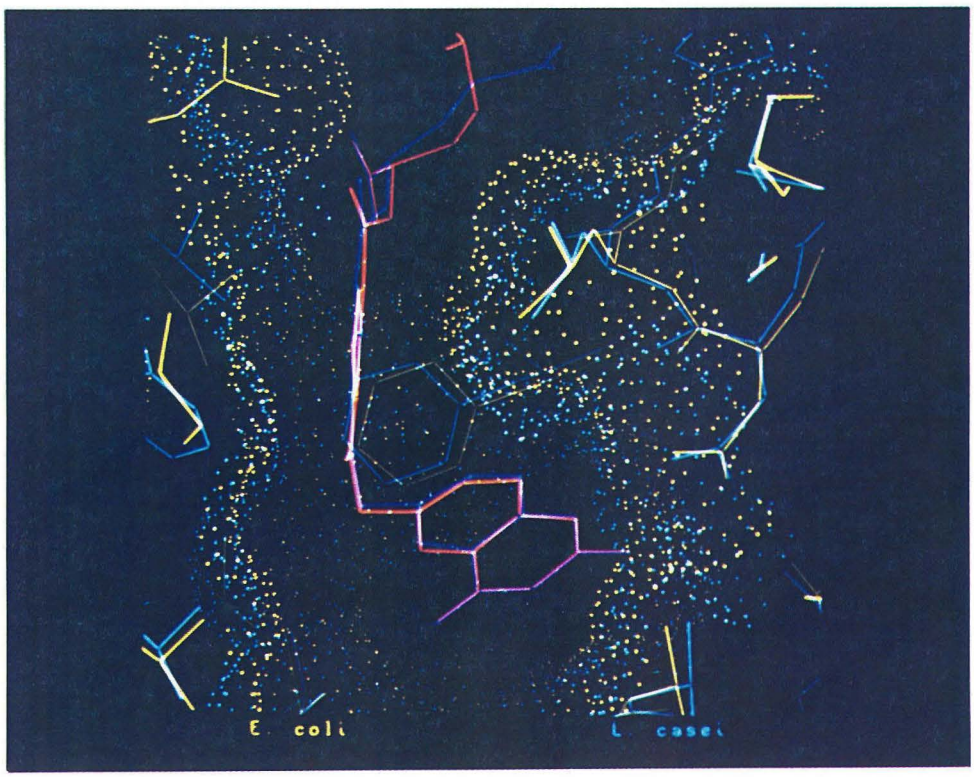


Figure 4.

Appendix IV

The text of this appendix is an article coauthored with Stephen J. Benkovic, Joseph A. Adams, and Carol A. Fierke. It has been submitted to *Pteridines* and is currently in press.

Insights into Enzymic Catalysis from Studies on Dihydrofolate Reductases

Stephen J. Benkovic[‡], Joseph A. Adams[‡,*], Carol A. Fierke[√] and Adel M. Naylor

Department of Chemistry
The Pennsylvania State University
University Park, Pennsylvania 16802

[‡] The Pennsylvania State University, to whom communication should be addressed.

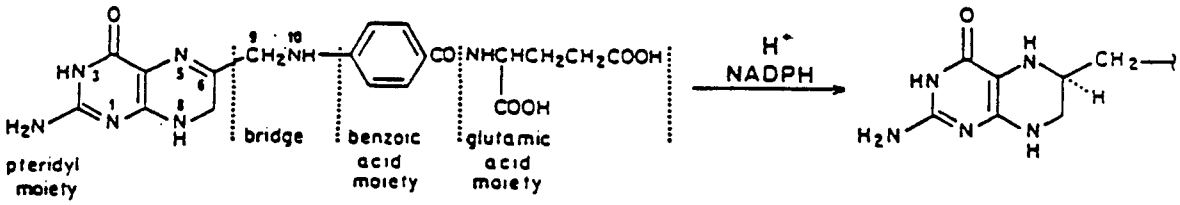
[‡] The Pennsylvania State University

[√] Duke University Medical Center

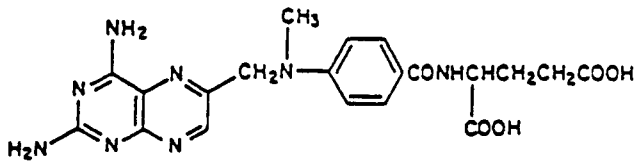
[*] California Institute of Technology

Introduction

Dihydrofolate reductase (5,6,7,8-tetrahydrofolate: $NADP^+$ oxidoreductase, EC1.5.1.3) catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate (H_2F)



H_2F H_4F



MTX (methotrexate)

to 5,6,7,8-tetrahydrofolate (H_4F). The enzyme is essential for both eucaryotes and procaryotes to avoid depletion of H_4F owing to its oxidation by thymidylate synthetase and thus directly influences pyrimidine biosynthesis. It has long been the target enzyme for chemotherapeutic agents and its inhibition by methotrexate (MTX) is mainly responsible for the mechanism of action of this drug (Blakley, 1985). Because of its biological and pharmacological importance, dihydrofolate reductase (DHFR) has been the subject of extensive structural and mechanistic investigations. In particular the X-ray crystallography has provided the structures of the MTX and MTX·NADPH enzyme

complexes for the E.coli and L.casei species respectively (Matthews et al., 1978 and 1979;Bolin et al., 1982; Filman et al., 1982). Stereochemistry has traced the course of the reduction and established that H_2F must be bound in a conformation in which the pteridine ring is rotated 180° relative to the C6-C9 bond of the side chain in MTX (Charlton et al., 1979). With the development of the protocols for site specific mutagenesis, it has become possible to pursue several novel lines of inquiry: i) to evaluate the role of hydrophobic residues in catalysis and in substrate and drug binding; ii) to ascertain the importance of proximal and distal residues on the acidity and function of the unique active site carboxylate; and iii) to determine the influence of multiple amino acid changes on the thermodynamic and kinetic characteristics of the active site by comparing the E.coli and L.casei enzymes.

Structural and Kinetic Features of Native DHFR

The active site structure of the E.coli DHFR·methotrexate·NADPH complex along with several conserved amino acid residues that have been altered by mutagenesis is depicted in Fig. 1. The NADPH has been positioned based on analogy to its observed locus in the L.casei complex. The active site is a cavity some 15 Å deep lined by hydrophobic side chains with the only acidic function present being the carboxyl of Asp-27. NADPH binds in an extended conformation with the nicotinamide moiety inserted through the entrance of the cavity and the pyrophosphoryl linkage making strong ionic and hydrogen bonding contacts (Arg-44, His-45). MTX, and by inference H_2F, lies towards the back of the site where the pteridine and benzoyl rings are surrounded by hydrophobic residues (e.g. Ile-5, Phe-31, Ile-50, and Leu-54). The Asp-27 interacts with the N2 and N3 nitrogens of the pyrimidine ring and does not make a direct hydrogen bonded contact with N5, despite the fact that protonation of N5 is

required to complete the reduction of the N5-C6 imine upon hydride transfer to C6.

The kinetic sequence for E.coli DHFR that predicts accurately the observed steady-state kinetic parameters and full-time course kinetics under a variety of substrate concentrations and as a function of pH was established by measuring the various ligand association and dissociation rate constants and pre-steady-state reaction transients employing stopped-flow fluorescence and absorbance spectroscopy (Fig. 2) (Fierke et al., 1987a). The enzyme has an intrinsic fluorescence that is quenched by the binding of ligands, particularly MTX. Independent measurements of the equilibrium binding for the various ligand combinations to form the binary and ternary reactant and product complexes gave dissociation constants in satisfactory agreement with those calculated from kinetic measurements of individual rate constants implying that the kinetic scheme is self consistent. Several features of the kinetic sequence germane to our considerations are: i) steady-state turnover is limited by H_4F release; ii) the nonequilibrium binding of H_2F and NADPH becomes ordered owing to the fact that the rate constant for H_4F dissociation is greatest from the mixed ternary E·NADPH·H_4F complex rather than either of the two possible product complexes; and iii) the overall reaction strongly favors H_4F formation ($K_{ov} \approx 10^{11}$) that is partially reflected in the internal equilibrium, E·NADPH·$H_2F \rightleftharpoons$ E·$NADP^+$·H_4F ($K_{int} \approx 10^3$).

The Active Site Asp-27

Reasoning from chemical precedent, the mechanism of imine reduction is expected to favor preprotonation at N5 followed by hydride transfer. However the absence of a proximal acid function calls into question this presumption and raises issues concerning the identity of the acidic residue and the rates of proton transfer steps to H_2F and medium.

Examination of the pH dependence of the steady-state turnover showed that the rate limiting step is H_4F release at pH <7.5 which changes to the chemical step of hydride transfer at pH >8.5. The pH-V_M transition is controlled by the ionization of an acidic residue with an apparent $pK_a \approx 8.1$ (Fig. 3). The rate constant for the isolated hydride transfer step is dependent on by an acidic active site group with a $pK_a \approx 6.5$, a value that is the same for all forms of the enzyme: free, binary and ternary complexes. Although binding of H_2F/H_4F are pH independent, the rate of dissociation of H_4F is significantly slower than the rate of hydride transfer up to pH 8.1 increasing the pK_a observed in V_M by ca. 1.6 pK_a units. The identity of the acidic active site residue was clearly shown to be the remote Asp-27 by its replacement with Asn or Ser resulting in two mutant enzyme forms that require preprotonated H_2F for activity (Howell, 1986).

Several lines of evidence from wild-type and mutant enzymes suggest that despite its cloistered location at the back of the active site cavity, the proton on Asp-27 readily exchanges with solvent namely: deuterium kinetic isotope effects on V_M/K_M (H_2F) and V_M that at high pH converge to the same value of ca. 3 (Chen, 1987); and the V_M/K_M (H_2F) pH-rate profile for the Tyr-31 mutant that exhibits a "hollow" in the vicinity of the apparent pK_a (Taira, 1987a). Similarly there is no sign that the proton transfer component of the chemical step becomes rate-limiting since i) the amplitude of the hydride transfer step in native DHFR, which is stoichiometric with enzyme, is not attenuated at a pH where the ternary complex exists equally in both protonated and nonprotonated forms and ii) the perturbed pK_a ($\Delta pK_a \approx 0$-2) of Asp-27 in various mutants does not correlate with the rate of hydride transfer. Thus, despite its location the carboxyl of Asp-27 can effectively function to donate its proton to N5 probably through a specific water channel

or via the periphery of the pterin ring(N3->O4->N5) (Gready, 1985). There is also evidence from mutants of DHFR that the carboxyl can be repositioned, e.g. by substitution of an Asp for Thr-113 and a neutral replacement at position-27. However, the lower catalytic efficiency [V_M/K_m (H_2F)] cannot be as yet unequivocally attributed to a decreased rate of proton transfer (Howell et al., 1987).

Pressing our line of inquiry further, we questioned whether the pK_a of Asp-27, and hence its interaction with H_2F and H_4F, might be coupled to events in the NADPH binding site in order to rationalize the NADPH assistance of H_4F dissociation. Mutation of Arg-44 to Leu-44 (Perry et al., 1987) that effectively deletes the ionic interaction of this side chain with the 2'-phosphate and pyrophosphate moiety of the nicotinamide cofactor in the binary complex as well as to neighboring residues in the free enzyme (e.g. Pro-66 and Ser-63) has a profound unexpected influence on the pK_a of Asp-27 in the free, binary, and ternary enzyme complexes. In all cases the pK_a is increased by ca. 2 pK_a units to 8.4 (Adams, 1987). Fig. 4 traces the amino acid backbone from Arg-44 in the N terminal direction to Asp-27 and in the C-terminal direction to Arg-52 and Arg-57--the latter residues form salt bridges to the two glutamate carboxylate residues of MTX. The distance from Arg-44 to Asp-27 is ca. 25 Å, suggesting subtle movement of the peptide chain is responsible for the pK_a change rather than an environmental effect. If the effect were purely electrostatic, a simple coulombic calculation ($\Delta pK_a = 244/Dr$ where D is the effective dielectric constant and r is the charge separation distance in Å) would require a dielectric constant of 5 (heptane ≈ 4). This is far less than the range of 50-100 reported for the influence of mutations on the pK_a of His-64 of subtilisin (Russell and Fersht, 1987). The Leu-44 mutant is also characterized by the absence of an accelerating effect by NADPH on the rate of

H_4F dissociation (2 s^{-1} vs. 12 s^{-1}) and a reduced rate for hydride transfer (50 s^{-1} and 950 s^{-1}), although the deuterium kinetic isotope effect (NADPD) on the hydride step remains the same (Adams, 1987). Thus we conjecture that the molecular origin of the NADPH effect commences at Arg-44 and extends to either or both Asp-27 and Arg-52/57 acting to weaken all three ionic/hydrogen bonding interactions with H_4F. The reduced rate of hydride transfer in this mutant, however, is not a result of an inadequate rate for proton transfer but is probably a consequence of alternate conformations for bound NADPH (Adams, 1987).

Hydrophobic Effects

In the study of designed mutant proteins whose purpose is to evaluate specific interactions, i.e. hydrogen bonds, ionic pairs or hydrophobic effects, a recurring question is their structural integrity versus the wild-type. Crystallographic studies on DHFR E.coli mutants (Asn-27 and Ser-27) showed that neither the MTX binding geometry nor the detailed three-dimensional topography of the enzyme was altered within ± 0.2 Å (Howell et al., 1986). However a substitution, be it a conservative deletion or replacement, will lead to conformational changes that may be localized or subtly propagated through the protein. Since in estimating the value of a side chain contribution to the free energy of binding the most relevant comparison is its deletion to a glycine, which because of its reduced volume, most certainly will be solvated differently at the active site. Thus in general $\Delta\Delta G_{obs}$ may not be equated to the binding energy of the moiety in question, although this term may be the predominant contribution (Fersht, 1986). With that caveat one can still assess the apparent binding energy and the relative contribution of a given residue to the active site ensemble.

It has been particularly instructive to examine the effect of a mutation (Phe-31 to Val-31) on the binding of the drug MTX. MTX is an example of a slow binding inhibitor where the bimolecular association proceeds in two stages via an initial encounter complex that isomerizes to the observed structure (Williams et al., 1979; Blakley, 1985). On the presumption that the free energies of the wild-type and mutant enzymes are equal, a free energy diagram has been constructed for the binding of MTX to the wild-type DHFR plus two mutants at Phe-31 (Taira et al., 1987b). Several salient results emerged: i) replacement of Phe-31 by Val-31 alters the accessibility of MTX to its binding pocket by reducing the bimolecular combination rate a hundred fold; ii) the value of edge to face aromatic-aromatic interactions estimated from the binding of MTX and 6,7-dimethylpteridine are ca. 3.0 kcal mol^{-1}. Since the salt bridge formed between N-1 and N-2 of MTX and Asp-27 may contribute ca. 1.8 kcal/mol^{-1} to the drugs binding affinity (Howell et al., 1986), hydrophobic interactions are at least as important determinants in the tight binding of MTX.

The more complex question of the relationship between substrate/product binding and enzymic catalysis can also be scrutinized by mutagenesis. For at least five mutants including those at Phe-31, Leu-54, and Thr-113 there is a linear free energy relationship between the rate constant for hydride transfer and the dissociation constant for H_2F plotted logarithmically (Benkovic et al., 1987). In essence there is ca. a 2 kcal mol^{-1} increase in the free energy of activation for the chemical step for every 1 kcal mol^{-1} decrease in the free energy of H_2F dissociation. In short the weaker ground state binding of the substrate translates into a slower rate of hydride transfer, an example of a loss in ground state as well as a greater loss in transition state stabilization (Jencks, 1975). The weaker binding of H_2F is also paralleled in

weaker H_4F binding so that the dissociation rate constant for this product no longer may be rate limiting.

Particularly informative is the Leu-54 to Gly-54 mutation, ca. 18 Å away from the locus of the hydride transfer, that decreases the rate of the chemical step from 950 s^{-1} to 14 s^{-1} and increases K_D (H_2F) from 0.21 μM to 140 μM (ca. 4.0 kcal/mol^{-1}) (Mayer et al., 1986). If the Gly substituent is viewed as too extreme a deletion, replacement of Leu-54 with Ile surprisingly produces an analogous change in the rate of the hydride transfer step (30 s^{-1}) (Murphy, 1987). The exquisite sensitivity of the locus to a single methyl group in the active site cannot at present be definitively rationalized; however an attractive hypothesis views bound H_2F as distributing between two conformers, one of which may be an unproductive complex akin to bound MTX whose equilibrium position is tuned by various residues. Moreover ab initio calculations suggest the pathway for hydride transfer to be exquisitely sensitive to the distance and angle between the participating atoms, i.e. changes on the order of less than 1 Å can decrease rates by factors of 10^2-10^3 (Wu and Houk, 1987).

Comparison of Mutiple Active Site Changes (E.coli and L.casei DHFR)

The detailed kinetic description of the E.coli DHFR catalyzed reaction provided an opportunity to evaluate the efficiency of this process. Given the definition of the kinetic sequence under physiological conditions (at concentrations for NADPH/$NADP^+$ and H_2F/H_4F approximating those of the E.coli cell) the turnover at steady state and its reciprocal the reaction flux were calculated (Fierke et al., 1987b). A maximum hypothetical catalytic turnover also can be computed after imposing H_2F diffusion to the E.NADPH complex (the predominant form of the enzyme in the cell) as the only rate-limiting step. At physiological conditions the E.coli DHFR would be operating at ca. 15%

efficiency, suggesting that the free energy cost of the various steps has been well optimized (Benkovic et al., 1987). Yet computer simulation revealed that there were a large number of pathways that efficiently solve the energetics of the reduction within the thermodynamic constraints imposed by the overall reaction equilibrium and the diffusional barrier of the enzyme/substrate combination. For example, the internal equilibrium constant ($E \cdot NADPH \cdot H_2F$ $E \cdot NADP^+ \cdot H_4F$) can be varied by 10^3 with less than a two-fold change in turnover maintaining the uniform binding constraint. Two related questions are prompted by the above calculations: i) is the catalytic process exhibited by the E.coli DHFR also employed by DHFRs from other species; ii) how many and what kinds of active site changes can be tolerated while still maintaining high catalytic efficiency?

The X-ray crystallographic structure of the active site amino acids of the E.coli and L.casei DHFRs were overlaid using MTX as a center for alignment resulting in a base ensemble of 40 amino acids within 5 Å of the drug. The active site equivalence was computed to be 55%; those amino acids that were either identical or have backbone interactions only were deemed equivalent. Despite these differences comparison of the total solvent accessible surfaces of the two proteins within this 5 Å radius showed them to be within 7% of each other and the protein surfaces to be remarkably congruent. However, several prominant changes (E.coli:L.casei) Ala-29: His-20; Leu-36: Val-35; Ile-50: Phe-49; and Arg-52: Lys-51 result in discernible differences in the positioning of the glutamate and benzoyl residues. A similar overlay was constructed for the NADPH binding site after positioning the NADPH in the E.coli site by analogy to its known position in the L.casei. In this case an active site ensemble of 58 amino acids 7 Å from the cofactor were counted with a 72% equivalence based on the above criteria. The close similarity in this section

of the active site surface between the two species is shown in Fig. 5 despite the amino acid differences and the fact that the NADPH was modeled into the E.coli structure without changes in side chain positions.

Elucidation of the kinetic sequence for the L.casei DHFR using the stopped-flow methods described earlier revealed the same features as reported for the E.coli (Benkovic, 1987; Andrews, 1988). Most importantly when the individual rate constants are expressed in terms of free energy the differences for each step (E.coli:L.casei) are within 1 kcal mol^{-1} equivalent to that of the E.coli enzyme. These free energy differences for the two enzymes are depicted for the bound species in Fig. 6.

It is apparent that an active site with multiple amino acid changes can result in similar kinetic sequences and catalytic efficiencies. Since single mutations can result in marked loss in catalytic activity, i.e. Leu-54 to Ile-54, it is reasonable to presume that some of these changes must be compensatory; others are probably neutral. One may hypothesize that the active site residues act as an ensemble to create a unique surface with the observed novel kinetic properties. The answer to the rhetorical question why this surface and not others may lie not only in the need for catalytic efficiency within the cell milieu but also in the stereochemical requirements of binding these substrates and in curtailing inhibition by others and the demands for stably folding a given primary sequence into this cavity shape. One may also speculate that residues strictly conserved during evolution may not be changed because no compensatory mutation exists that satisfies all these three requirements within the cell.

References

Adams, J. A. (1987) unpublished results.

Albery, W. J., & Knowles, J. R. (1976) Biochemistry 15, 5631-5640.

Andrews, J. (1988) unpublished results.

Benkovic, S. J., Fierke, C. A., & Naylor, A. M. (1987) Science, in press.

Blakley, R. L. (1985) in Folates and Pterins pp. 191-253, Wiley, New York.

Blakley, R. L., & Cocco, L. (1985) Biochemistry 24, 4772-4777.

Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., & Kraut, J. (1982) J. Biol. Chem. 257, 13650-13662.

Charlton, P. A., Young, D. W., Birdsall, B., Feeney, J., & Roberts, G. C. K. (1979) J. Chem. Soc. Chem. Commun. 922-924.

Chen, J. -T., Taira, K., Tu, C. D., & Benkovic, S. J. (1987) Biochemistry 26, 4093-4100.

Fersht, A. R. (1987a) Trends Biochem. Sci. 12, 301-304.

Fierke, C. A., Johnson, K. A., & Benkovic, S. J. (1987b) Biochemistry 26, 4085-4092.

Fierke, C. A., Kuchta, R. D., Johnson, K. A., & Benkovic, S.J. (1987) Cold Spring Harbor Symp. on Quant. Biol. 52, in press.

Filman, D. J., Bolin, J. T., Matthews, D. A., & Kraut, J. (1982) J. Biol. Chem. 257, 13663-13672.

Gready, J. E. (1985) Biochemistry 24, 4761-4766.

Howell, E. E., Villafranca, J. E., Warren, M. S., Oatley, S. J., & Kraut, J. (1986) Science 231, 1123-1128.

Howell, E. E., Warren, M. S., Booth, C. L. J., Villafranca, J. E., & Kraut, J. (1987) Biochemistry, in press.

Jencks, W. P. (1975) Advances in Enzymology 43, 219-410.

Matthews, D. A., Alden, R. A., Bolin, J. T., Filman, D. J., Freer, S. T., Hamlin, R., Hol, W. G. J., Kisliuk, R. L., Pastore, E. J., Plante, L. T., Xuong, N. -H., & Kraut, J. (1978) J. Biol. Chem. 253, 6946-6954.

Matthews, D. A., Alden, R. A., Freer, S. T., Xuong, N., & Kraut, J. (1979) J. Biol. Chem. 254, 4144-4154.

Mayer, R. J., Chen, J. -T., Taira, K., Fierke, C. A., & Benkovic, S. J. (1986) Proc. Natl. Acad. Sci. 83, 7718-7720.

Murphy, D. J. (1987), unpublished results.

Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittleman, M. S., Matthews, C. R., Chen, J. -T., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E., & Kraut, J. (1987) Biochemistry 26, 2674-2682.

Russel, A. J., & Fersht, A. R. (1987) Nature 328, 496-500.

Taira, K., Chen, J. -T., Fierke, C. A., & Benkovic, S. J. (1987a) Bull. Chem. Soc. Jpn. 60, 3025-3030.

Taira, K., & Benkovic, S. J. (1987b) J. Med. Chem., in press.

Williams, J. W., Morrison, J. F., & Duggleby, R. G. (1979) Biochemistry 18, 2567-2573.

Wu, Y. D., & Houk, K. N. (1987), personal communication.

Acknowledgements

We thank W.A. Goddard III for use of the computer graphics facility for graphical display.

Footnote

Abbreviations: DHFR, dihydrofolate reductase; H_2F, 7,8-dihydrofolate; H_4F 5,6,7,8-tetrahydrofolate; MTX, methotrexate.

BIOGRAF is an interactive molecular simulation 3-dimensional graphics program from BIODESIGN, Inc., Pasadena, CA.

Figure Legends

Fig. 1. Carbon backbone structure of E.coli DHFR·MTX·NADPH and the side chains for key amino acids depicted using BIOGRAF*.

Fig. 2. Kinetic scheme for DHFR at 25°C, MTEN buffer, N, NADP; NH, NADPH; H_2F, dihydrofolate; H_4F, tetrahydrofolate.

Fig. 3. Observed rate constants for hydride transfer (-··-), H_4F dissociation (---) and V_M (-) as a function of pH.

Fig. 4. Amino acid structure of E.coli from Arg-44 to Arg-57 (C-terminal direction) and Asp-27 (N-terminal direction) depicted using BIOGRAF*.

Fig. 5. Superpositioning of the active sites of the E.coli (gold) and L.casei (blue) DHFRs with NADPH as a center for alignment. All residues within 7 Å were included. Some key residues that are visible in the left hand panel (adenine portion) are (E.coli:L.casei); Gln-65: Gln-65 (top); Arg-44:Arg-43 (middle right); Arg-98:Gln-101 (bottom); and Ser-64:His-64 (middle left). In the right hand panel (nicotinamide portion) the key residues are (E.coli:L.casei); Ile-94:Ala-97 (upper right); Phe-31:Phe-30 (middle right); Asp-27:Asp-26 (lower right); and Ile-50:Phe-29 (upper left). These residues are depicted using BIOGRAF*.

Fig. 6. Plot of $\Delta\Delta G$ (E.coli minus L.casei) calculated from equilibrium constants vs. various bound species that are identified in Fig. 2. E[TS]* represents the transition state for the chemical hydride transfer step calculated from the forward rate constant. All free energy values are pH independent and are measured using standard state concentrations of 1 M NADPH and H_2F. T = 25°C.

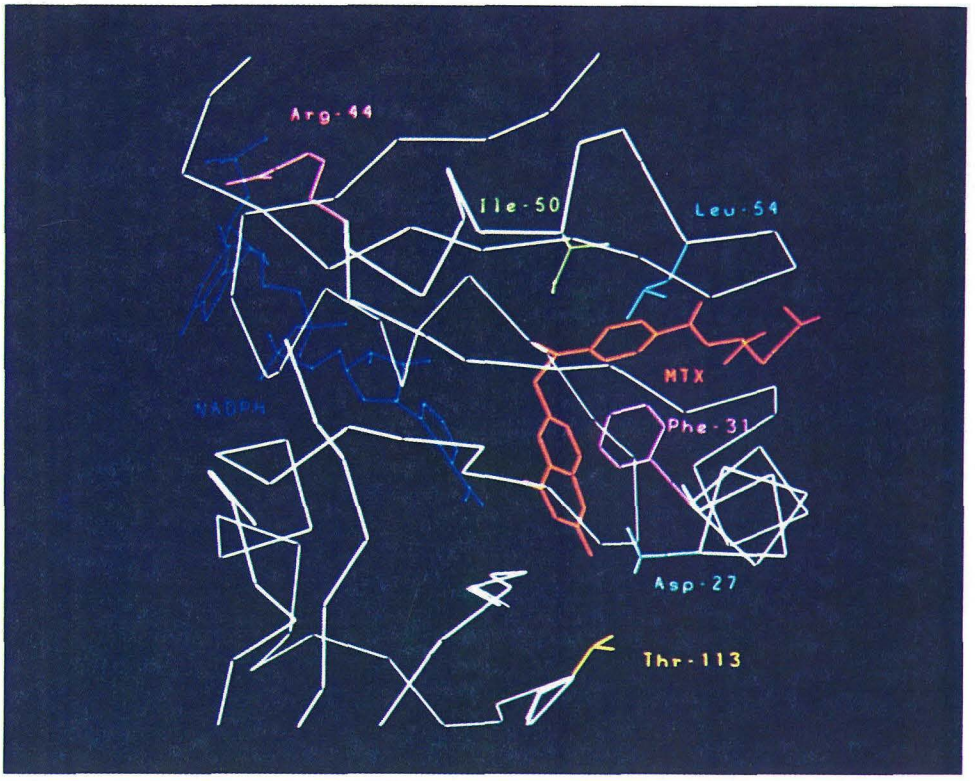


Figure 1.

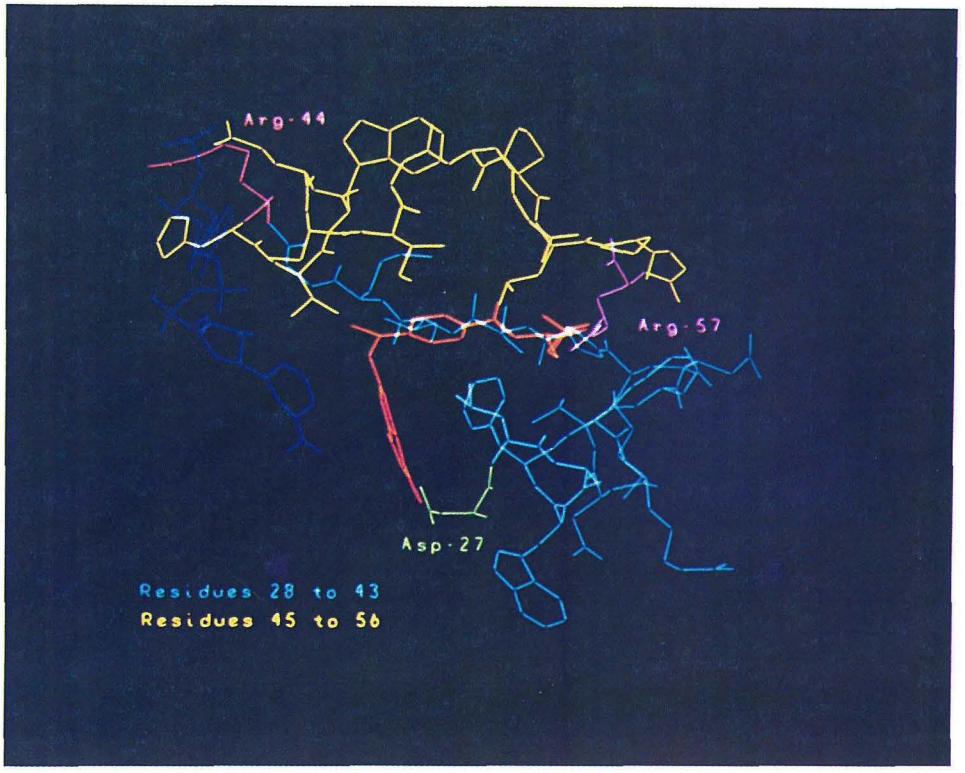


Figure 4.

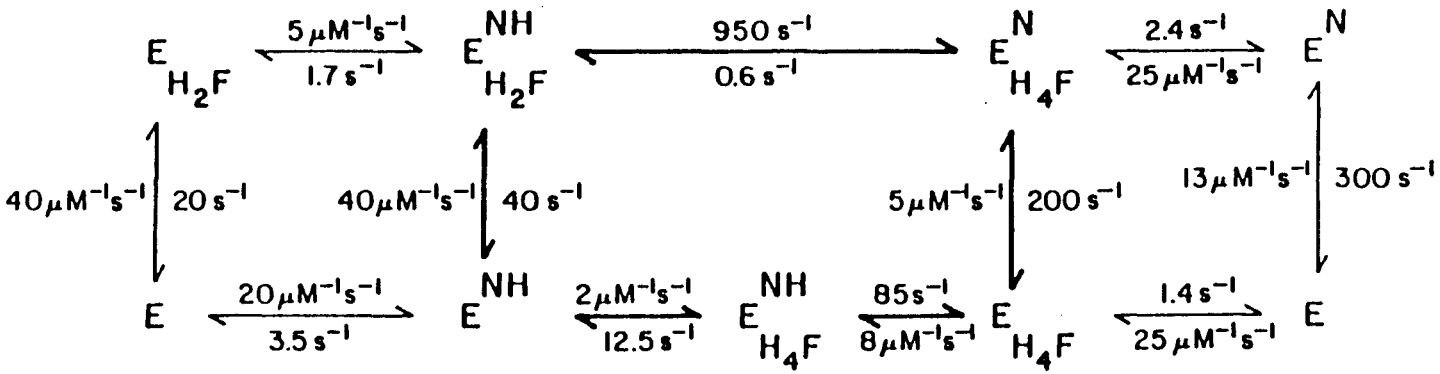

E H2F
5 μM⁻¹s⁻¹
1.7 s⁻¹
E NH H2F
950 s⁻¹
0.6 s⁻¹
E N H4F
2.4 s⁻¹
25 μM⁻¹s⁻¹
E N
40 μM⁻¹s⁻¹
20 s⁻¹
40 μM⁻¹s⁻¹
40 s⁻¹
5 μM⁻¹s⁻¹
200 s⁻¹
13 μM⁻¹s⁻¹
300 s⁻¹
E
20 μM⁻¹s⁻¹
3.5 s⁻¹
E NH
2 μM⁻¹s⁻¹
12.5 s⁻¹
E NH H4F
85 s⁻¹
8 μM⁻¹s⁻¹
E H4F
1.4 s⁻¹
25 μM⁻¹s⁻¹
E

Figure 2.

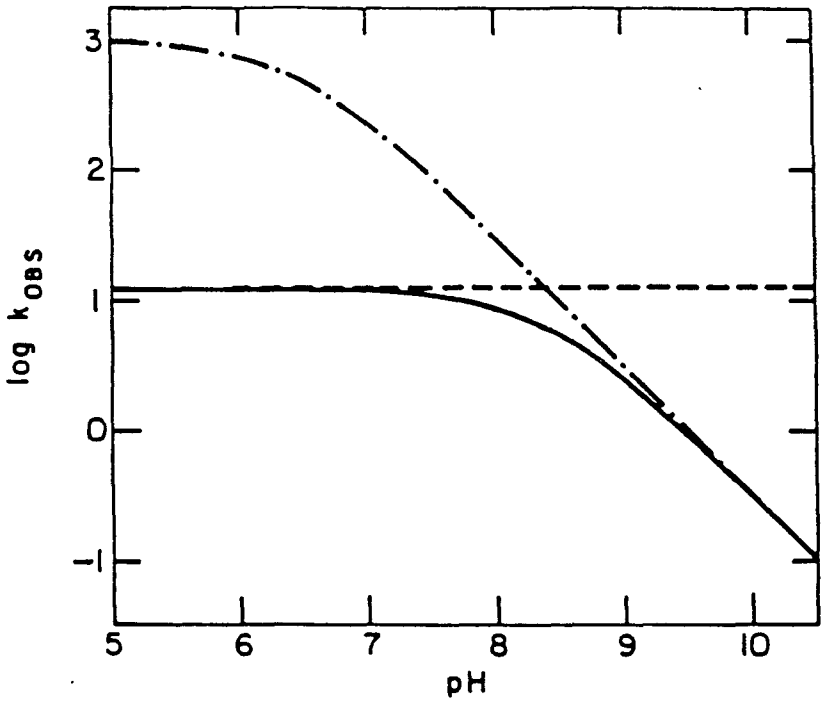


Figure 3.

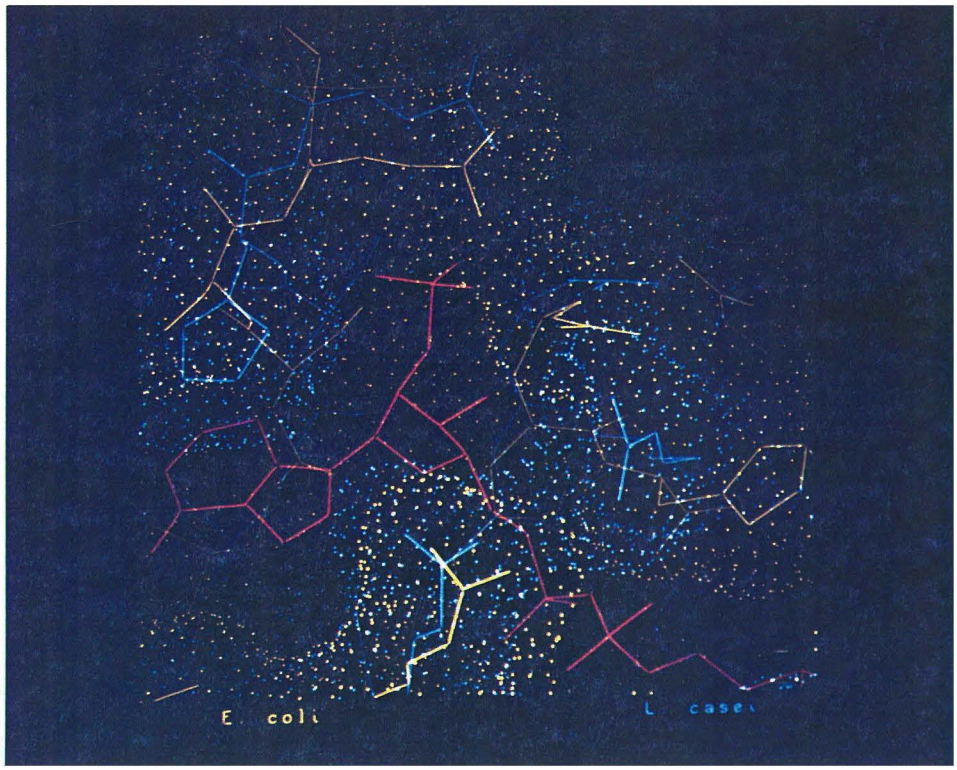


Figure 5. left

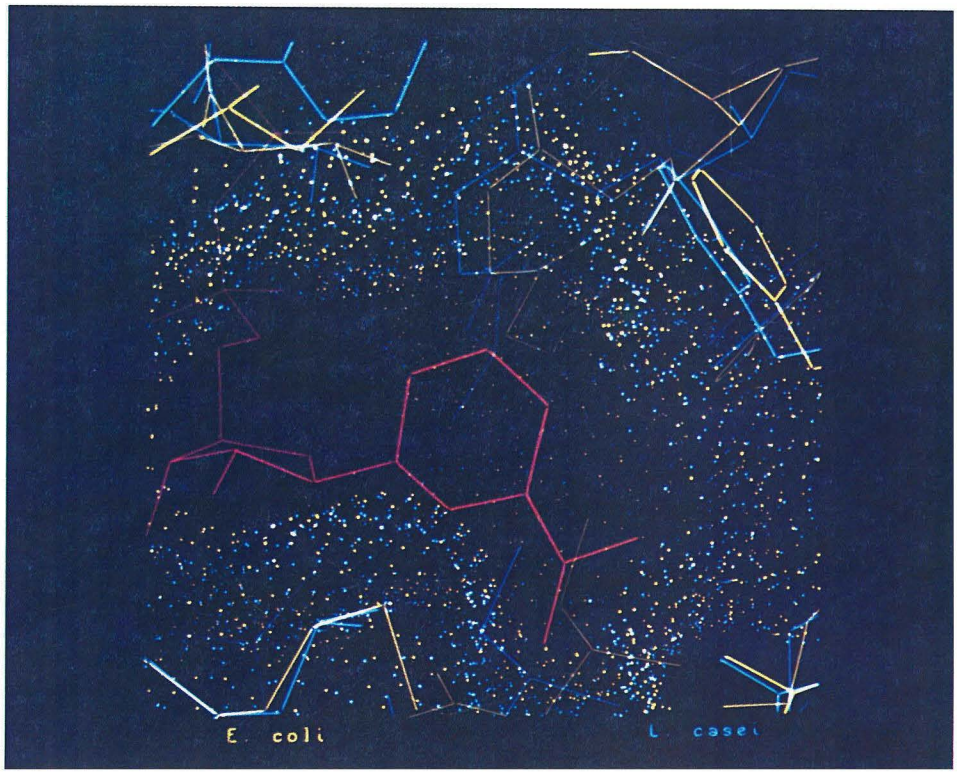


Figure 5. right

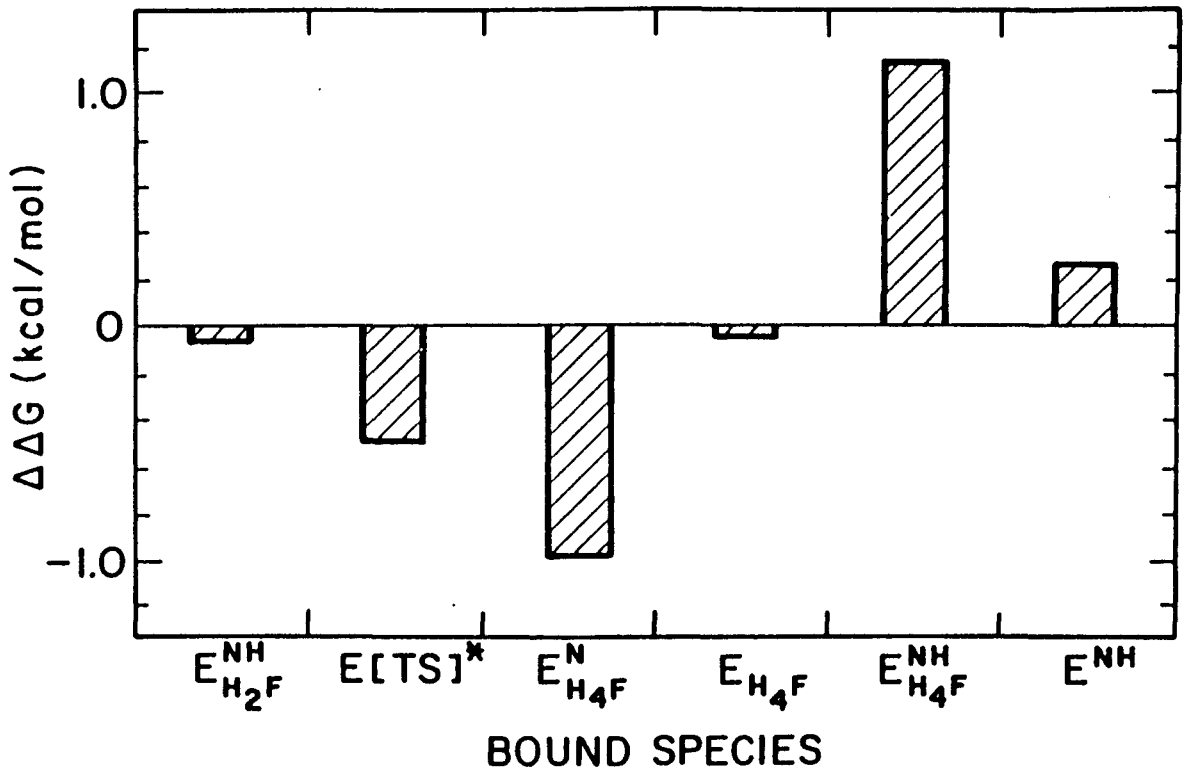

1.0
0
-1.0
ΔΔG (kcal/mol)
$E^{NH}_{H_2F}$
E[TS]*
$E^{N}_{H_4F}$
E_{H_4F}
$E^{NH}_{H_4F}$
E^{NH}
BOUND SPECIES

Figure 6.

Appendix V

The solvent potential used for the molecular simulations is derived from the Jorgensen transferable, intermolecular, potential model[1]. This scheme is a pairwise additive site-site potential with the interaction sites positioned at the nuclear centers. The form of the intermolecular potential function is given by Equation AV-1.

$$U^{(2)}(\mathbf{r_1},\mathbf{r_2}) = \sum_{i,j}^{\text{all sites}} \frac{A_{ij}}{r_{ij}^{12}} + \frac{C_{ij}}{r_{ij}^{6}} + \frac{q_i q_j e^2}{r_{ij}} \qquad (AV-1)$$

Here, $\mathbf{r_i}$ is the set of vectors defining the positions of the atoms in two different molecules, and $r_{i,j}$ is the distance between center i and center j. For the TIPS model,[1] the constants have the values A_{OO} = 580,000 kcal Å^{12}/mol, C_{OO} = -525.0 kcal Å^6/mol, and q_H = 0.40.

The intramolecular potential is given by Equation AV-2.

$$U^{(1)}(\mathbf{r}) = \frac{1}{2}a((r_{OH_1} - d_{OH_1})^2 + (r_{OH_2} - d_{OH_2})^2) +$$

$$\frac{1}{2}b(\theta_{HOH} - t_{HOH})^2 + \frac{1}{2}c(r_{HH} - d_{HH})^2 \qquad (AV-2)$$

Here, r_{ij} is the distance between atoms i and j within the single molecule, θ_{HOH} is the angle between the two OH bonds, d_{ij} and t_{ij} are the equilibrium distance and angle for the isolated molecule, respectively. The TIPS model uses the experimental geometry for water. Thus, θ_{HOH} = 104.5° and r_{OH} = 0.96 Å. For the Pettitt flexible modification[2], the constants have the values a = 1059.2 kcal/mol/Å^2, b = 68.1 kcal/mol/radian^2, and c = 76.5 kcal/mol/Å^2.

References

(1) W. L. Jorgensen. *J. Amer. Chem. Soc.*, **1981**, *103*, 335.

(2) L. X. Dang, B. M. Pettitt. *J. Phys. Chem.*, **1987**, *91*, 3349.

Well, Bob, I won't have to spend
the rest of my life wondering ...